University of Tennessee, Knoxville
TRACE: Tennessee Research and Creative Exchange

Doctoral Dissertations | Graduate School

8-2001

Advances in modeling techniques for aqueous organic molecules at hydrothermal conditions : searching for clues as to the chemical origins of life

Eric Martin Yezdimer
University of Tennessee

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

Recommended Citation

Yezdimer, Eric Martin, "Advances in modeling techniques for aqueous organic molecules at hydrothermal conditions : searching for clues as to the chemical origins of life. " PhD diss., University of Tennessee, 2001.
https://trace.tennessee.edu/utk_graddiss/6470

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Eric Martin Yezdimer entitled "Advances in modeling techniques for aqueous organic molecules at hydrothermal conditions : searching for clues as to the chemical origins of life." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Chemistry.

Peter Cummings, Major Professor

We have read this dissertation and recommend its acceptance:

Ariel Chialvo, Robert Compton, Jeffery Kovac

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Eric Martin Yezdimer entitled "Advances in modeling techniques for aqueous organic molecules at hydrothermal conditions: searching for clues as to the chemical origins of life." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Chemistry.

Peter Cummings, Major Professor

We have read this dissertation and recommend its acceptance:

Ariel Chialvo

Robert Compton

Jeffrey Kovac

Accepted for the Council:

Interim Vice Provost and
Dean of The Graduate School

ADVANCES IN MODELING TECHNIQUES FOR AQUEOUS ORGANIC MOLECULES AT HYDROTHERMAL CONDITIONS: SEARCHING FOR CLUES AS TO THE CHEMICAL ORIGINS OF LIFE

A Dissertation
Presented for the
Doctor of Philosophy
The University of Tennessee, Knoxville

Eric Martin Yezdimer
August 2001

Copyright ® Eric Martin Yezdimer, 2001

All rights reserved

For Kristina

ACKNOWLEDGMENTS

I would like to thank the following people, without whose assistance this achievement would not have been possible: Andrei Sharygin, Caibin Xiao, Andrei Plyasunov, Wenbin Liu, Douglas Doren, Ariel Chialvo, Peter Cummings, Dave Frurip, Thomas Driesner, Mike Simonson, and Hung-Chih Li. I must especially thank Shinichi Sakane, Josef Sedlbauer, and Robert Wood, all of whom I have had the great pleasure of closely working with over the past several years and too whom I cannot properly express my heart felt gratitude. I would also like to thank my mother, father and brother for always being supportive in all of my endeavors. To my best friend Erin Monaco, I thank her for always helping me to see the bright side of things. Finally, I must thank my soon-to-be wife, Kristina Clark, who during the darkest days of graduate school constantly reminded me that finishing was indeed worth the effort.

ABSTRACT

This dissertation seeks to address some of the basic problems involved in modeling the thermodynamics of infinitely dilute organic molecules at hydrothermal conditions and presents an equation of state for simple acyclic organic solutes. The proposed equation should allow accurate predictions to 630 K, an increase of roughly 150 degrees over earlier work. However, several of the predicted thermodynamic trends were unexpected. Free energy of formation calculations strongly suggested that at high temperature ($T > 480$ K) the existence of larger, more complex organic molecules was preferred over the existence of smaller, less complex organic molecules. These results are quite opposite to the situation at room temperature, where the existence of smaller organic solutes is preferred over large molecules. Similar trends were also found in the predicted organic aqueous solubilities. As an independent check of the free energy predictions, classical molecular dynamics calculations were conducted and the results were found to be in good agreement with the equation of state's predicted trends. Generally speaking, our results suggest that a hydrothermal medium is considerably more favorable for the chemical evolution of complex organic molecules (both in terms of the free energy of formation and aqueous solubility of a particular organic solute) than an aqueous solution at ambient conditions. These findings offer addition support for the high-temperature origin of life hypothesis. In addition, a novel new computational technique was developed, based solely on first principles, which should allow for relatively fast and accurate free energy predictions of bulk aqueous organics, helping to further eliminate some of the undesirable guess work out of empirically based thermodynamic modeling methods.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABC-FEP	*Ab-initio*/Classical Free Energy Perturbation
EOS	Equation of State
HKF	Helgeson, Kirkham and Flowers Equation of State
LJ	Lennard-Jones
MC	Monte Carlo Molecular Simulation
MD	Molecular Dynamics Simulations
NDIS	Neutron Diffraction with Isotope Substitution
TI	Thermodynamic Integration
SCF	Supercritical Fluid
QM/MM	Quantum Mechanical Molecular Mechanics

PART 1

LIFE AT HYDROTHERMAL CONDITIONS

1.1 Introduction

One of the fundamental questions in biochemistry is how life evolved from a general collection of inorganic molecules into the vast arrays of interlocking structures, reaction pathways, proteins, and DNA molecules that are so commonplace in even the most basic biological cell. So far we have just scratched the surface of the chemistry involved in the functioning of the simplest single-cell organisms. Moreover, the difficulty in understanding how these processes initially began is compounded by the sobering fact that life on Earth has had over 4 billion years of evolution to effectively mask its humble beginnings. This dissertation has been undertaken with the purpose of further exploring some of the basic chemical issues and computational problems relating to the initial emergence of life, in particular the prediction of aqueous chemical equilibrium. Our hope is that this work will serve to further stimulate future discussions in these and related areas, eventually leading to a broadening of our general understanding and conceptions about the nature of life.

Some of the most interesting pioneering work done on the origin of life, focused around gaining a clearer understanding of how microbial life initial evolved on Earth. Using ribosomal RNA sequencing, Woese (1987) attempted to create a family tree for contemporary single cell organisms. By examining the similarity of each organism's genetic sequencing, Woese was able to make a rough estimation of the relative evolutional age of several different microbial organisms. However, it is still not possible to determine the absolute timescale between two different microbial evolution paths. A general outline of the microbial family tree is given in Figure 1.1 (Woese, 1987; Pace, 1997). While the exact classification of the root microbial organism (the so-called "Eve"

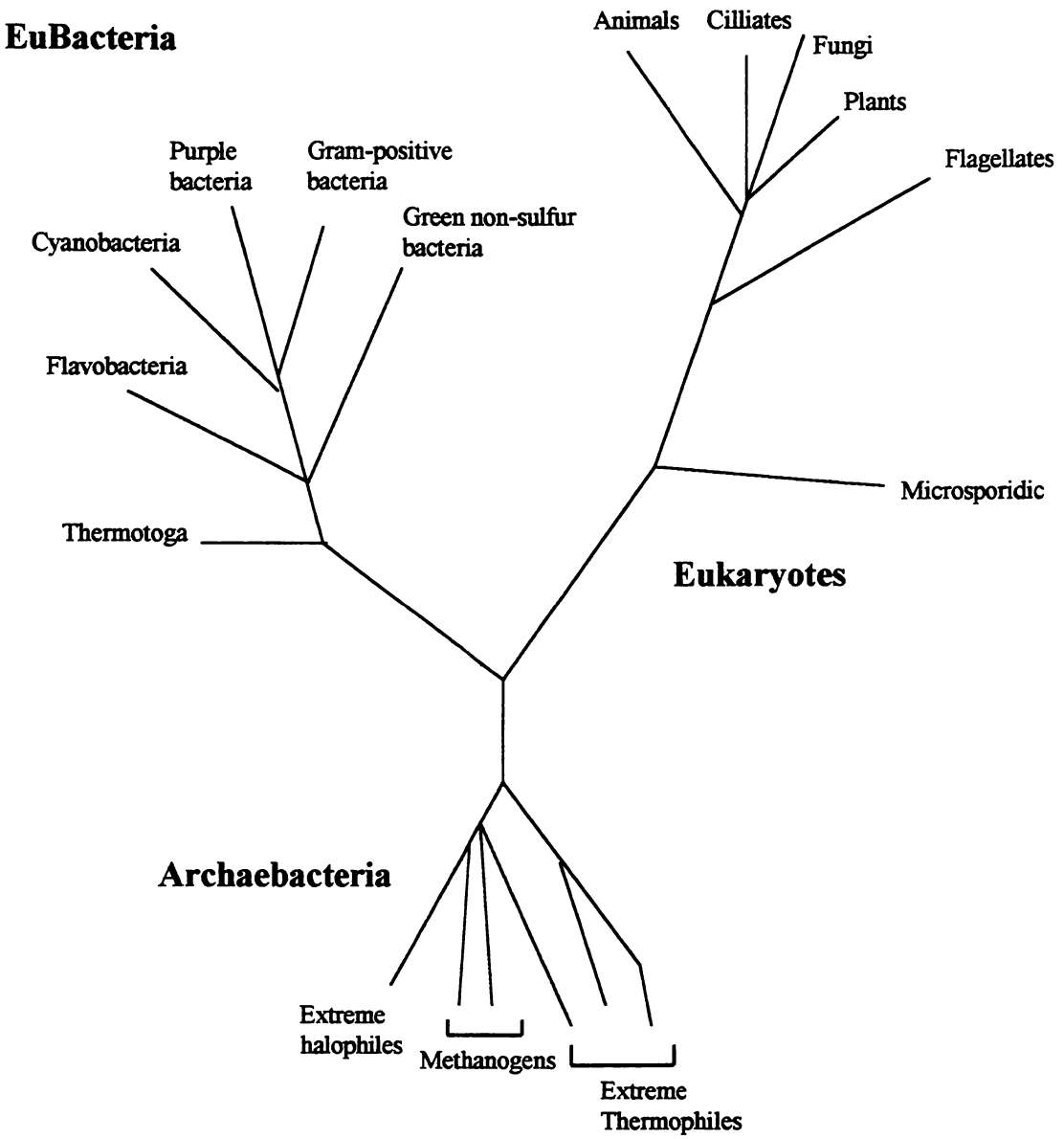

Figure 1.1 : Universal Phylogenetic Tree of Woese (1987). The distances between the groups were determined by comparing rRNA sequences. As their name suggests archaebacteria are thought to be some of the most primitive organisms on the planet.

microbe) is still unclear, most researchers believe that the first microbial organisms originated close to the archaea line. However, some other phylogenetic studies have suggested that the ancestral root organism(s) actually originated on the early bacterial line (Doolittle and Brown, 1994). There appears to be a general consensus that the evolution of mitochondrion and chloroplasts came relatively late in the evolutionary history of single cell organisms and that chemolithoautorophs and/or archaeabacteria are considerably more primitive than phototrophic or organotrophic cells (Pace, 1997).

Besides their atypical energy sources, most archaebacteria are classified as extremeophiles. As the name implies, extremeophiles prefer to live in environments that are normally lethal to most other forms of bacteria. Two main sub-classifications of extremeophiles are the halophiles and thermophiles. Halophiles reach their peak biological activity at very high salt concentrations, anywhere from 20% to saturation (Yamauchi and Kinoshita, 1993; Christian and Waltho, 1962). Thermophiles, on the other hand, require high solution temperatures (333 – 373 K) to reach their peak biological activity. Still yet a third group, hypothermophiles, require exceptionally high temperatures (>373 K) in order to function properly. To date the highest optimum growth temperature at which a bacteria sample has been successfully cultured is between 383 – 393 K. Daniel (1992) and Segerer *et al.* (1993) have speculated that some bacteria forms may be sustainable at 423 K, and Cragg and Parkes (1994) have reported possible evidence for biological process occurring at 442 K. Unfortunately the current experimental bacterial culturing techniques are somewhat crude at high temperatures and

pressures and it is quite possible that even more extreme hypothermophiles will be successfully grown and identified under laboratory conditions as the technology becomes better refined.

Prior to the late 1970's it was thought that it was impossible for life to exist at temperatures above 328 K. This narrow view of the Earth's available biosphere changed suddenly in 1977 when Corless *et al.* (1979) reported the discovery of the first deep-sea hydrothermal vent ecosystem. This largely self-contained ecosystem was found to function in the complete absence of solar radiation and derived its base source of energy from chemolithoautoroph bacteria. The areas near the escaping vent fluid were found to be rich in theromophilic bacteria. Many of these microbes were later found to make their living through oxidation/reduction reactions of inorganic gases and other surrounding minerals according to the reaction,

$$CO_2 + 4\,H_2 \rightarrow 2\,H_2O + CH_4, \qquad (1.1)$$

(Stevens and McKinley, 1995). Later studies found these ecosystems to be widespread through out the world's oceans and that they were extremely productive, generating an estimated 10^{10} kilograms of biomass annually (McCollom and Shock, 1997).

Up until the early 80's it was commonly believed that the beginnings of life on Earth had its origin in astrochemistry, the early atmosphere or tidal pools. The proposed mechanisms for the creation of basic organic molecules tended to focus on various intermittent high-energy sources (*i.e.*, lighting, meteorite impacts). While it is clear from carbon arc experiments that the required energy and chemical pathways existed in those scenarios in order to create organic matter form inorganic precursors (*i.e.*, CO_2, CO, HCN, *etc.*) , there was no suggested physical mechanism to concentrate the organics in an

environment that would be conducive to further biochemical development. In addition, some of the oldest bacteria fossil remains have been dated at 3.9 billion years old (compared to an estimated 5 billion year old Earth), implying that the emergence of life must have occurred relativity quickly after the Earth began to cool. Organic molecules present on/near the planet's primordial surface or atmosphere would also be under the constant impact of solar ultraviolet radiation, making the evolution of microorganisms difficult. Considering all these factors, it seemed that the emergence of life was somewhat of a fluke accident, whose probability of it occurring would be all but inconceivable.

The discovery of deep-sea hydrothermal ecosystems, along with the groundbreaking phylogenectic work of Woese soon allowed for an alterative, deep sea/subsurface origin of life theory to take shape. Thermophilic organisms, in addition to being some of the most primitive on the planet, are able to thrive in environments very similar to those thought to be present on the abiotic, early Earth. The escaping vent fluid temperature can range up to 623 K (with even higher temperatures possible in the oceanic crust's subsurface), thus supplying the energy required to form various complex organic molecules. Modern thermophiles have shown their ability to derive energy from very basic inorganic sources (Eqn. (1.1) and other similar reactions). Molecular hydrogen can be generated easily by abiotic mechanisms that involve the reaction of water with iron-bearing basalts (which comprise a significant proportion of the Earth's oceanic crust) (Pace, 1997) or by degassing of igneous rocks in hydrothermal fluids (McCollom and Shock, 1997). The mixing of hydrothermal fluids with the surrounding cooler seawater near geothermal vents also creates disequilibriums in the oxidation/reduction reactions of

sulfur, carbon, iron, and manganese containing compounds which chemolithotropic organisms could possibility exploit (Chen and Morris, 1972; Mottl *et al.*, 1979; Seyfried and Mott, 1982; Welhan, 1988; Shock 1990, 1992; Lillet *et al.*, 1993). There is also mounting evidence that chemolithotropic organisms may be quite abundant deep in the Earth's subsurface (Gold, 1992; Fredrickson and Onstott, 1996, Lovely and Chapelle, 1995; Stevens and McKinley, 1995). Since these environments occur in areas where they were shielded from the sun's destructive high-energy radiation, photo-degradation would not have been a concern. These observations coupled with the wide and frequent natural distribution of hydrothermal vent systems all over the world, offered a view of evolution that was quite different from the earlier prevailing theories. Instead of appearing to be the result of chance chemical encounters, the deep sea/subsurface origin theory portrayed life as emerging as part of a common natural process which could readily occur on any other solid planetary body possessing liquid water.

Other areas of interest with thermophilic bacteria include natural petroleum extraction and industrial chemical synthesis. The average temperature and pressure inside the sub-surface tends to increase linearly with depth (~25 K per kilometer and 30 MPa per kilometer). Despite our current limited knowledge of archaebacteria, large numbers of microbial organisms could easily be living at depths up to 6 km. In fact many petroleum fields have reported the existence of microscopic organisms in their well fluids. The existence of microbial colonies at these depths could have profound consequences on our understanding of the dynamics and replenishing/depletion of natural hydrocarbon reservoirs. Enhanced secondary oil recovery technologies and processes have been and are currently being developed to exploit the mineral consuming

thermophiles to increase the amount of extractable petroleum (Yen, 1990; Donaldson, 1991). Even though many field trials have suggested the possible economic advantage of these methods, they have still not yet obtained wide industrial acceptance. Extremophiles are also finding increasing use in industrial chemical reactors and genetic engineers have imagined designing "super bugs" which could possibly replace traditional industrial catalysts or perform very highly specialized reactions.

1.2 Organic Synthesis

The "hot water" theory of life offered a sharp contrast to earlier theories in that the emergence of life appeared to be the result of a constant process rather than the result of infrequent and intermittent events. However, the "hot water" theory is far from complete and several important questions still remain. For instance, the exact types, concentrations, and half-lives of organic molecules generated in hydrothermal systems are still highly debatable. Obviously if biochemically important compounds like proteins, lipids, ATP, NAD, *etc.* do not posses a significant chemical life span or thermodynamically favorable free energies of formation in a particular hydrothermal environment, it would be very difficult for modern bacteria and/or primitive single celled organisms to survive/evolve under those conditions. Numerous experiments have been conducted to examine aqueous organic synthesis/decomposition reactions, however these experiments have yielded mixed results. Lavrent'yev *et al.* (1985), White (1984), Vallentyne, (1968, 1964), Miller and Bada (1988) and Bada *et al.* (1991) have found that amino acids and various proteins quickly decompose at elevated temperatures. Miller and Bada concluded that microorganisms were not likely to be capable of surviving at

temperatures significantly above 393 K. However these conclusions were soon criticized by Shock (1992a) who argued that since hydrothermal environments contain a significant amount of dissolved salts and minerals the aforementioned experiments were not representative of natural hydrothermal systems and were instead only indicative of a purely hydrous environment. Shock also stressed the importance of properly considering the effects of any catalyzes, buffering, or protective mechanisms that may be present in the natural systems (Table 1.1). In addition, large concentrations of CO_2 and H_2 (as well as other gases) are common to hydrothermal systems and would most likely aid in forcing the organic/inorganic equilibrium back toward the organic side (Shock, 1992b).

Other studies employing more realistic hydrothermal systems have met with more success. French (1971) showed that ethanol, acetic acid, propanoic acid, butanoic acid, and various ketones could be synthesized from siderite ($FeCO_3$) at hydrothermal conditions. This result was quite unexpected and was discovered only after French noticed an organic smell coming from his apparatus. Oró and Han (1966) showed that significant concentrations of various aromatic compounds could be generated from methane and silica gel at 1273 K. Still yet, Mukhin (1976) while attempting to experimentally model the chemistry of an "underwater volcano" was able to synthesize aldehydes, amino acids, aromatic hydrocarbons, and polycyclic aromatic hydrocarbons at temperatures greater than 1173 K. However the residence time in the reactor was quite small, less than 1 second. In this sense Mukhin's "underwater volcano" was very similar to carbon arc experiments, implying that underwater volcanism and other geothermal activity may have had a significant role in the initial inorganic generation of the first

Table 1.1: Average chemical composition of a typical hydrothermal vent fluid[a]. All concentrations are given in millimolal.

Property / Concentration	Vent Fluid	Average Seawater
T	623 K	275 K
pH	4.54	7.8
O_2 (aq)	0	0.076
H_2 (aq)	0.45	0
H_2S(total)	7.5	0
Na^+	439	464
Ca^{2+}	16.6	10.2
Mg^{2+}	0	52.7
Ba^{2+}	0.01	0.00014
Al^{3+}	0.0047	0.00002
Fe	0.750	0.0000015
Cl^-	496	541
ΣCO_2[b]	5.72	2.3
SiO_2(aq)	17.3	0.16
K^+	23.2	9.8
Mn^{2+}	0.70	0
Cu^{2+}	0.01	0.000007
Pb^{2+}	0.000194	0
Zn^{2+}	0.089	0.00001
SO_4^{-2}	0.6	27.9

[a] Taken from Shock and Schulte (1998). Also see Von Damm, 1990; Von Damm *et al.*, 1985; Welhan and Craig, 1983.

[b] $\Sigma CO_2 = CO_2(aq) + HCO_3^- + CO_3^{-2}$

basic organic molecules on Earth. In a effort to more precisely model a hydrothermal system, Yanagawa and Kobyashi (1992) used a complex mixture of various salts (2mM $Fe(NH_4)_2(SO_4)_2$, 0.6 mM $MnCl_2$, 0.1 mM $ZnCl_2$, 0.1 mM $CuCl_2$, 20 mM $CaCl_2$, 0.1 mM $BaCl_2$, 50 mM NH_4Cl) and were able to synthesize several amino acids from methane and carbon dioxide.

From a thermodynamic modeling perspective, one method to accurately model these complex aqueous systems is through an equation of state for all the electrolyte and non-electrolyte species present in a given hydrothermal solution. The most notable equation of state for infinite dilute aqueous solutes is the revised-HKF equation (Helgeson *et al.*, 1981; Shock and Helgeson, 1988,1990). Using this equation Shock and Schulte (1998) have modeled the formation of several acyclic alkene, alcohols, ketones, and organic acids under several different possible early Earth geochemical models. Their results indicated that there was a large enough thermodynamic driving force to allow for the significant formation of acyclic organic compounds. An earlier study also predicted that there was a strong thermodynamic driving force for initial peptide polymerization (Shock, 1992c).

Despite its great success, it has been shown that the revised-HFK equation has difficulty predicting the general behavior of infinite dilution partial molar volume data for CH_4, CO_2, H_2S, NH_3, and H_3BO_3 near the critical region of pure water (O'Connell *et al.*, 1996). New equations based on Kirkwood's fluctuation solution theory, which can describe the correct thermodynamic trends in the near critical and critical regions, are showing themselves to be simpler and more accurate (Sedlbauer *et al*, 2000). Using a functional group approach, Yezdimer *et al.* (2000) have been able to parameterize this

equation of state to give very accurate predictions for a very wide range of aqueous organic molecular solutes (to temperatures of 600K and pressures of 100 MPa). Amend and Helgeson (1997) have also recently performed similar functional group parameterization of the HKF equation; however, that parameterization was submitted before the latest series of high temperature experimental data became available (Criss and Wood, 1996; Inglese and Wood, 1996; Inglese *et al.*, 1996; 1997) and as a result their equation was not parameterized with any data above 400 K.

The difficulty in drawing concrete conclusions about the reactivity in hydrothermal systems is that a hydrothermal vent is never in equilibrium. The movement of vent fluid from the subsurface, to vent, to seawater, back to the subsurface is an open process and at best the system can only be considered as existing at a steady state. Even if the "true" organic/inorganic equilibrium did strongly favor the inorganic side, significant accumulations of organic materials in the surrounding rocks and clays is very possible through a natural application of Le Châtelier's law. This is not to say that the equilibrium thermodynamic modeling and experimental studies of organic formation do not yield significant insight into the probability of various organic molecules being created. However, one is often forced to treat these studies as rough organic molecular inventories, from which complex biological structures could have emerged. They also help to provide general estimates of what temperatures and pressures modern life may persist too. It is also worth noting that beside the implications in origin of life theories, thermodynamic modeling of aqueous organic synthesis/decomposition reactions are also of great importance in developing more environmentally friendly industrial solvents (*i.e.*,

water based solvents) and designing more effective trace hazardous organic waste disposal techniques (*i.e.*, supercritical water oxidation).

1.3 Peptide and Proteins

Besides the formation of basic organic molecules, modern cells require the formation of biopolymers in order to synthesize enzymes. Normal peptide polymerization proceeds through a condensation reaction and, as expected, production of large molecular weight polymers is not favorable in a purely hydrous environment. In hydrothermal systems this is not necessarily the case, since the presence of various metal ions and complexes may effectively reduce the water activity and help catalyze the reaction(s). There are several experimental studies of note. The early work of Oró and Guidry (1961) reported that the polymerization of glycine was possible in aqueous solution at temperatures from 403 to 433 K. Similar results were reported by Lowe *et al.* (1963) for the polymerization of amino acids at 363 K. Later studies showed that the polymerization of glycine could be increased by increasing the NaCl + Cu(II) concentration (Rode and Schwendinger, 1990). However the lengths of the peptide chains generated were quite small, normally less than 6 monomer units and are on the order that one might expect for an aqueous condensation reaction.

This is not to say that large peptide synthesis in water is impossible, as modern cells are able to routinely build proteins in an aqueous media (even at temperatures at least as high as 383 K). The secret lies in using the correct enzymes or catalysts. Using a glass reaction tube Yanagawa and Kobyashi (1992) were able to synthesize peptide-like polymers of molecular weight from 1000 – 2000 Da at temperatures of 523 K from

aqueous solutions glycine, L-alanine, L- valine, and L-aspartic acid. However, the resulting polymers were not true peptide chains in the classic sense and IR spectrums suggested the presence of Si-O-Si bonds in the peptide chains, further highlighting the potential role of metal catalysts in hydrothermal systems. It is also important to note the observation of Cody (2000), that many modern protein make extensive use of transition metal sulfides and that these species are very commonly found in hydrothermal solutions. This situation allows us to speculate that perhaps the first proto-proteins were inorganic metal or organometalic catalysts, which later chemicaly evolved the extensive use of amino acids chains.

Another interesting, less explored, issue raised by Miller and Bada (1988) revolved around the problem of maintaining molecular chirality at elevated temperatures. Several experimental studies have been done on measuring water solvated amino acid racemization rates (Bada and Man, 1980; Bada, 1985a,b; Boehm and Bada, 1984; Wehniller, 1984; Bada and Miller, 1987; Smith and Sivakua, 1983). These studies, however, have been limited to temperatures between 353 – 403 K for the basic amino acids (*i.e.* Ala, Val, Ile, and Leu). Some initial work by Miller and Bada (1988) estimated the free amino racemization rates to be on the order of hours at 523 K, suggesting that it would be nearly impossible for organisms living at those temperatures to maintain their molecular chirality, although their estimations were based on simple Arrhenius extrapolations. Moreover disagreement among various experiential studies can range up to tens of kJ/mol in the calculated activation energies and from up to 1 to 2 orders of magnitude in the measured rate constants. These discrepancies could easily amount to calculated protein life times differing anywhere from hours to weeks at a particular state

point and are not suitable for making accurate predictions of protein efficiency and functionality at high temperatures.

1.4 Outline of Recent Work

As seen from the above discussion, the overall experimental results as to whether or not various organic molecules are stable at hydrothermal conditions trends to be confusing and unclear. In the final analysis, the difficulty in obtaining a clearer understanding lies in the vast complexity of the systems of interest. For significant future progress to made, a step-by-step breakdown of the chemistry involved in hydrothermal systems will most likely be necessary. Theoretical approaches offer considerable promise in this respect because they allow the evaluation of a single reaction or chemical species at one time. Predictions of more complex systems can then be constructed based on rigorous and standardized chemical concepts. In this dissertation we present some of the basic theoretical approaches to answering these questions and attempt to help lay the ground work for future theoretical investigations of more complex hydrothermal systems.

In the next part, we outline the semi-empirical equation of state developed by Yezdimer *et al.* (2000) for infinite dilute aqueous organics. As discussed earlier accurate equations of state for aqueous organics are absolutely essential to thermodynamically model organic synthesis/decomposition reactions at hydrothermal conditions. Molecular dynamic computer simulations are then conducted in part 4 in an attempt to independently confirm the equation of state's near critical thermodynamic predictions for alkane solubilies in water.

Part 3 mainly focuses on the evaluation of several different types of classical polarizable potential forcefield models. Both the structure and thermodynamics of four different polarizable models are examined and compared to experimental results. The polar character of liquid water is highly dependent on the temperature and density of the system (*i.e.*, the dipole moment of water in the ideal gas state at ambient conditions is 1.85 D while it is estimated to range from 2.6-2.9 D in the liquid phase). Since the accurate portrayal of liquid water's changing polar character with temperature and density is thought to play a significant role in the thermodynamic modeling of aqueous hydrophobic organic molecules and membrane/protein assemblies, it was imperative to select a classical water model that most accurately modeled real water's polarization. The pure component vapor-liquid phase envelope of the most accurate classical water model was then examined and used to make a direct connection between molecular dynamics simulations and experimental results through the law of corresponding states.

As will be seen the following parts, our underlining approach to the prediction of the thermodynamic properties of aqueous organics is through a description of the free energy of hydration. Although this method appears very accurate, it is normally dependent on the availability of partial molar volume and solubility data. Since our interest lies mainly at non-ambient conditions, where data is scarce or difficult to obtain, this presents a problem. Ideally we would like to employ molecular dynamics as an accurate and cheaper alterative to direct experimentation. However as will be seen in part 4, while molecular dynamics simulations can provide qualitatively accurate thermodynamic trends (*i.e.*, free energies) as a function of temperature and pressure, quantitative agreement with the known experimental values is difficult to obtain. The

origin of this problem lies mainly with the necessary use of an empirically derived molecular potential surface. In an effort to devise a more accurate method of parameterizing these models, part 5 deals with the development of a new *ab-initio* quantum mechanical method for determining the necessary classical molecular force fields.

In the final section of this dissertation, part 6, the material presented in parts 1-5 is briefly reviewed. Using the EOS presented in part 2, the thermodynamic stability of several different classes of aqueous organic molecules is examined. The overall tend appears to be that larger organic solutes are actually energetically preferred over that of their smaller homogenous analogs at high temperatures. This behavior is exactly the opposite of that at near ambient conditions, where the existence of small organic solutes is preferred over the existence of large organic solutes. These findings tend to support the hydrothermal origin of life theory and indicate that hydrothermal solutions represent a unique, not to mention a very favorable, solvent environment for large and complex organic molecules.

References

Amend, J. P. and Helgeson, H. C. 1997. *Geochim. Cosmochim. Acta*, **61**, 11.

Bada, J.L. 1985a. *Ann. Rev. Earth Planet. Sci.*, **13**, 241.

Bada, J. L. 1985b. *Chemistry and Biochemistry of the amino acids: Racemization of amino acid* ed. G. Barrett. Chapman and Hall, London, 397.

Bada, J. L. and Man, E. H. 1980. *Earth Sci. Rev.*, **16**, 21.

Bada, J.L. and Miller, S.L. 1987. *BioSystems*, **20**, 21.

Boehm, M. and Bada, J.L. 1984. *Proc. Natl. Acad. Sci.*, **81**, 5263.

Chen, K.Y. and Morris, J.C. 1972. *Environ. Sci. Technol.*, **6**, 529.

Christian, J. H. B. and Waltho, J. A. 1962. *Biochim. Biophys. Acta*, **65**, 506.

Cragg, B. A. and Parks, R. J. 1994. *Proc. Ocean Drilling Program, Sci. Res.*, **139**, 509.

Criss, C. M. and Wood, R. H. 1996. *J. Chem. Thermodyn.*, **28**, 723.

Cody, G. III. ; Boctor, N.; Hazen, R. and Yoder, H. Jr. 2000. "Experimental investigations into the potential hydrothermal origins of life." *2000 International Chemical Congress of Pacific Basin Societies*. Honolulu, Hawaii.

Corliss, J.B.; Dymond, J.; Gordon, L.I.; Edmond, J.M.; von Herzen, R.P. ; Ballard R.D.; Green, K.; Williams, D.; Bainbridge, A. ; Crane, K. and van Andel, T. 1979. *Science*, **203**, 1073.

Daniel, R. M. 1992. *Orig. Life. Evol. Biosphere*, **22**, 33.

Donaldson, E.C. (ed.) 1991. *Microbial enhancement of oil recovery- Recent Advances*, Elsevier Science, Oxford.

Doolittle, W.F. and Brown, J.R. 1994. *Proc. Natl. Acad. Sci.*, **91**, 6721.

Fredrickson, J.K. and Onstott, T.C. 1996. *Sci. Am.*, **275**, 68.

French, B. M. 1971. *Amer. J. Sci.*, **272**, 37.

Gold, T. 1992. *Proc. Natl. Acad. Sci.*, **89**, 6045.

Helgeson, H. C.; Kirkham, D. H. and Flowers, G. C. 1981. *Am. J. Sci.*, **281**, 1249.

Inglese, A. and Wood, R. H. 1996. *J. Chem. Thermodyn.*, **28**, 1059.

Inglese, A. ; Sedlbauer, J. and Wood, R. H. 1996. *J. Sol. Chem.*, **25**, 849.

Inglese, A. ; Sedlbauer, J.; Yezdimer, E. M. and Wood, R.H. 1997. *J. Chem. Thermodyn.*, **29**, 517.

Lavrent'yev, G. A. 1985. *Geochem. Int.* **22**, 52.

Lillet, M.D.; Butterfield, D.A.; Olson E.J.; Lupton J.E.; Macko, S.A. and McDuff R.E. 1993. *Nature*, **364**, 45.

Lovely, D. and Chapelle, F. 1995. *Rev. Geophys.*, **33**, 365.

Lowe, C.U.; Rees, M.W. and Markham, R. 1963. *Nature*, **199**, 219.

McCollom, T.M. and Shock, E.L. 1997. *Geochim. Cosmochim. Acta,* **61**, 4375.

Miller, S. L. and Bada, J. L. 1988. *Nature*, **334**, 609.

Mottl, M. J.; Holland, H.D. and Corr, R.F. 1979. *Geochim. Cosmochim. Acta*, **43**, 869.

Mukhin, L.M. 1976. *Orig. Life Evol. Biosphere,* **7**, 355.

O'Connell, J.; Sharygin, A.V. and Wood, R. H. 1996. *Ind.Eng. Chem. Res.*, **35**, 2808.

Oró, J. and Han, J. 1966. *Science,* **153**, 1393.

Oró, J. and Guidry, C. L. 1961. *Arch. Biochem. Biophys.*, **93**, 166.

Pace, N. R. 1997. *Science,* **276**, 734.

Rode, B. M. and Schwendinger, M. G. 1990. *Orig. Life Evol. Biosphere*, **20**, 401.

Sedlbauer, J.; O'Connell, J.P. and Wood, R.H. 2000. *Chem. Geol.*, **163**, 43.

Segerer, A.H.; Burggraf, S ; Fiala, G. ; Huber, G. ; Huber, R; Pley, U and Stetter, K.O. 1993. *Orig. Life Evol. Biosphere*, **23**, 77.

Seyfried, W., E., Jr. and Mottl., M., J. 1982. *Geochim. Cosmochim Acta*, **46**, 985.

Shock, E.L. 1990. *Orig. Life Evol. Biosphere,* **20**, 331.

Shock, E. L. 1992a. *Orig. Life. Evol. Biosphere,* **22**, 67.

Shock, E. L. 1992b. *Orig. Life Evol. Biosphere*, **22**, 135.

Shock, E. L. 1992c. *Geochim. Cosmochim. Acta*, **56**, 3481.

Shock, E. L. and Helgeson, H.C. 1988. *Geochim. Cosmochim. Acta*, **52**, 2009.

Shock, E. L. and Helgeson, H.C. 1990. *Geochim. Cosmochim. Acta*, **54**, 915.

Shock, E. L. and Schulte, M. D. 1998. *J. Geophys Res.*, **103**, 28513.

Smith, G. G. and Sivakua, T. 1983. *J. Org. Chem.*, **48**, 627.

Stevens, T.O. and McKinley, J.P. 1995. *Science,* **270**, 450.

Vallentyne, J. R. 1964. *Geochim. Cosmochim. Acta*, **28**, 157.

Vallentyne, J. R. 1968. *Geochim. Cosmochim. Acta*, **32**, 1353.

Von Damm, K. L. 1990. *Ann. Rev. Earth. Planet. Sci.*, **18**, 173.

Von Damm, K. L.; Edmond, J. M. and Grant B. 1985. *Geochim. Cosmochim. Acta* , **49**, 2197.

Wehniller, J. F., 1984. *Quat. Res.* (NY) **22**, 109.

Welhan, J.A. 1988. *Chem. Geol.*, **71**, 183.

Welhan J. A. and Crag H. 1983. *Hydrothermal Processes at Seafloor Spreading Centers* ed. P.A. Rona *et al.*, Plenum Press.

White, R. H. 1984. *Nature*, **310**, 430.

Woese, C. R. 1987. *Microbiological Reviews*, **51**, 221.

Yamauchi, K. and Kinoshita, M. 1993. *Prog. Polym. Sci.*, **18**, 763.

Yanagawa, H. and Kobayashi 1992. *Orig. Life Evol. Biosphere*, **22**, 147.

Yen, T. F. 1990. *Microbial enhanced oil recovery: Principle and Practice,* CRC Press Boca Raton, Florida.

Yezdimer, E. M.; Sedlbauer, J. and Wood, R. H. 2000. *Chem. Geol.*, **164**, 259.

PART 2

DEVELOPMENT OF AN EQUATION OF STATE FOR INFINITELY DILUTE AQUEOUS ORGANIC SOLUTIONS AT HIGH TEMPERATURE AND PRESSURE

2.1 Background

In recent years there has been an increasing interest in the ability to predict thermodynamic properties of chemical reactions of aqueous organic species at high temperatures and pressures by geochemists and chemical engineers. For example, the formation and transformation of hydrocarbon deposits, the study of deep sea hydrothermal ecosystems, power plant chemistry and the potential for the decomposition of hazardous organic waste all require accurate knowledge of the thermodynamic properties of aqueous organic solutes at high temperatures and pressures. However, measuring these thermodynamic properties for every organic compound of interest is not practical. What is needed is a method of predicting properties of many compounds from experiments on a few key compounds.

It is well established that functional group additivity near room temperature allows reasonably accurate predictions of thermodynamic properties of aqueous solutes in the standard state of infinite dilution (Cabani *et al.*, 1981; Hoiland, 1986; Gianni and Lepori, 1996). Several recent studies (Criss and Wood, 1996; Inglese, *et al.*, 1997) have shown the accuracy of functional group additivity for infinitely dilute partial molar volumes and heat capacities, V_2^0 and $C_{p,2}^0$, up to temperatures of 523 K at 28 MPa. Using the revised HFK equation of state (Shock and Helgeson, 1988, 1990; Tanger and Helgeson, 1988; Shock *et al.*, 1989, 1992), Amend and Helgeson (1997a) also demonstrated that the assumption of group additivity holds at above standard conditions and derived equation of state coefficients for several functional groups. However, their

findings were based only on earlier experimental results which represent just a part of the temperature and pressure range of the latest experimental results (*e.g.* Criss and Wood, 1996; Inglese *et al.*, 1996a, 1996b, 1996c, 1997; Xiao *et al.*, 1997).

This part presents a functional group approach to the predictions of standard thermodynamic properties of aqueous solutions of organic solutes at high temperatures and pressures. Criss and Wood (1996) and Inglese *et al.* (1997) have given equations for functional group contributions to V_2^0 and $C_{P,2}^0$ at temperatures to 523 K along the isobar of 28 MPa, but their equations cannot be extrapolated to higher temperatures and are limited to a single isobar. In this part we pursue the following goals. First we present a functional group scheme which takes into account the latest experimental results for V_2^0 and $C_{P,2}^0$. Then, we predict the standard thermodynamic properties for a wide variety of organic species for which there are little or no experimental results. Finally we use an equation which allows reasonable extrapolations of the standard properties of these organic species to higher temperatures and pressures. For this purpose the model recently presented by Sedlbauer *et al.* (2000) was used. In addition, a full account of these issues have been given in the recent publication (Yezdimer *et al.*, 2000).

2.2 Experimental Literature Review

An extensive literature search for partial molar volumes and partial molar heat capacities of aqueous organic electrolyte and nonelectrolyte solutions at infinite dilution has been conducted. Although the amount of experimental data available in the literature for these properties at temperatures above 353 K and pressures higher than 0.1 MPa is

still rather limited, a range of new experimental results appeared mainly in the recent years. The databases contain over 2800 and 900 data points, respectively, and are available for free via the Internet at http://people.vslib.cz/josef.sedlbauer /download.htm. However, due to constraints that will be discussed later, only about 930 data points for V_2^0 and 260 data points for $C_{P,2}^0$ were used in this study. A brief description of the available high temperature and pressure volumetric and heat capacity data is given below for each major class of compounds that has been included in this study.

2.2.1 Hydrocarbons

Virtually no standard molar volume or standard molar heat capacity data on aqueous hydrocarbons at infinite diluation were found above 298 K, except for methane, which have been reported to 725 K by Hnedkovsky *et al.* (1996, 1997).

2.2.2 Alcohols

Standard molar volumes for simple straight chain alcohols such as methanol, ethanol, propanol, butanol, and pentanol to 348 K have been reported (Jolicoeur and Lacroix, 1976; Alexander, 1959; Makhatadze and Privalov, 1989; Hoiland, 1986; Hoiland and Vikingstad, 1976; Nakajima *et al.*, 1975; Sakurai, Nakamura, and Nitta, 1994). Properties for several branched alcohols were also measured by Sakurai (1987, 1988, 1989). Data for 1-pentanol have been reported (Inglese *et al.*, 1996 a) to a temperature of 413 K and pressure of 19 MPa. Experimental results for 1-propanol, 1,4-butanediol, and 1,6-hexanediol (to 523 K) along the 28 MPa isobar have been reported

by Criss and Wood (1996). In addition, Xiao *et al.* (1997) reported partial molar volumes of methanol to a temperature of 573 K and a pressure of 13.5 MPa. Friedman and Scheraga (1965) conducted measurements on various alcohols to 323 K, however their findings are not consistent with those of other studies and their results were not used in this study. The standard molar heat capacities of 1-propanol, 1,4-butanediol, and 1,6-hexanediol have been reported by Inglese and Wood (1996b) to 523 K at 28 MPa.

2.2.3 Carboxylic Acids

The amount of experimental data on the partial molar volume of carboxylic acids at temperatures above 298 K is quite limited. Makhatadze, Medvedkin, and Privalov (1990) and Allred and Wooley (1981) have reported data to 348 K for acetic and propanoic acids. The only high temperature volumetric data (to 523 K and at 28 MPa) on carboxylic acids above 373 K were reported by Criss and Wood (1996) for propanoic acid, adipic acid, and succinic acid. Partial molar heat capacities to 525 K for acetic, propanoic, and succinic acids have been reported by Inglese *et al.* (1996c). Heat capacities for acetic and propanoic acids to 393 K were measured by Ackermann and Schreiner, and to 348 K by Makhatadze and Privalov (1990).

2.2.4 Amines and Amides

Several laboratories (Makhatadze, Medvedkin, and Privalov, 1990; Hoiland, 1986; Kiyohara, Perron, and Desnoyers, 1975; Kaulgud, Bhagde, and Shrivastava, 1982) have reported data on acetamide, propionamide, and butylamine to 348 K. Also, monoethanolamine has been measured by Maham *et al.* (1994) to a temperature of 353 K.

Criss and Wood (1996) have reported volumetric results for 1,4-butanediamine, 1-propylamine, and propionamide to 523 K and at 28 MPa. Data for the partial molar heat capacity of 1-propylamine, 1,4-butanediamine, and 1,6-hexanediamide have been reported to 523 K and at 28 MPa by Inglese *et al.* (1997).

2.2.5 Amino Acids

Standard molar volumes of the L-isomers of glycine, alanine, aspartic acid, glutamic acid, glutamine, isoleucine, leucine, serine, threonine, arginine, proline, asparagine, and valine have been reported to 328 K by several authors (Kikuchi *et al.*, 1995; Chalikian *et al.*, 1993; Hakin *et al.*, 1994a, 1994b, 1995, 1997; Yasuda *et al.*, 1998; Duke *et al.*, 1994). Also data on NH_3^+-$(CH_2)_n$-COO^- where $4 \leq n \leq 8$ and 11-aminoundecanoic acid have been reported to 328 K by Chalikian *et al.* (1993). The standard molar heat capacities for asparagine, glutamine, glutamic acid, valine, leucine, isoleucine, glycine, alanine, serine, threonine, arginine, and proline have been measured to 328 K (Hakin *et al.*, 1994a, 1994b, 1995, 1997; Duke *et al.*, 1994). Standard molar volumes and standard molar heat capacities were reported for glycine and alanine to 318 K and 100 MPa by Chalikian *et al.* (1994).

2.2.6 Electrolytes

Standard molar volumes for several sodium carboxylates and ammonium chlorides have been reported by Sakurai (1973), Sakurai *et al.* (1975a, 1975b) and Allred and Wooley (1981) to 318 K. For organic electrolytes at high temperatures, only partial molar volumes of sodium propanoate, sodium acetate, sodium benzenesulfonate and

propylamine hydrochloride to 523 K at 28 MPa have been reported by Criss and Wood (1996). Measurements on standard molar heat capacities for the same compounds have been reported to 523 K at 28 MPa by Inglese *et al.* (1997).

2.2.7 Other Properties

Available experimental results on other standard partial molar thermodynamic properties are generally more scarce in the literature than for the standard molar volumes and standard molar heat capacities. We collected smaller databases on standard molar compressibilities, κ_2^0, hydration enthalpies, $\Delta_{hyd}H_2^0$, and hydration Gibbs energies, $\Delta_{hyd}G_2^0$, which include about 70, 180 and 200 data points, respectively, of which 60, 50 and 90 data points were used in our calculations. Standard molar compressibilities of hydrocarbons at 298 K and alcohols from 278 K to 313 K have been reported by Hoiland (1980, 1986). Standard molar hydration enthalpies and hydration Gibbs energies at 298 K were based on a comprehensive compilation of Cabani *et al.* (1981), which was completed with data for hydrocarbons to 313 K in case of hydration enthalpies (Dec and Gill, 1983; Naghibi *et al.*, 1987) and to 353 K in case of hydration Gibbs energies (Wilhelm *et al.*, 1977). These databases are also available on the Internet site mentioned above.

2.3 Equation of State

In a functional group scheme, the general equation for any standard molar thermodynamic property of an ion or neutral solute is given by,

$$\Xi_2^0 = (1-z)\cdot(\Xi_{SS} + \Lambda_P) + \sum_{i=1}^{N} n_i \cdot \Xi_{2,i}^0, \qquad (2.1)$$

where N is the total number of functional groups present in a given compound, n_i is the number of occurrences of each specific functional group in a given compound, Ξ_2^0 is the thermodynamic property of a compound, $\Xi_{2,i}^0$ is the property contribution to each functional group, and z is the charge of the particle. Ξ_{SS} in Eqn.(2.1) is the standard state term, which is the property of a point mass (Ben Naim, 1987). Since one Ξ_{SS} is needed for each particle (ionic or neutral) and we wish to remain consistent with the hydrogen convention scale for aqueous ions, which requires $\Xi_2^0(H^+) = 0$, a *(1-z)* multiplication term has been introduced to Eqn.(2.1). The standard state properties are easily calculated from the equation $G_{SS} = R\cdot T\cdot \ln[\lambda^3 / V_m]$ by appropriate differentiation, where V_m is the molar volume in the standard state and λ is the de Broglie wavelength of the particle. The standard state terms are; 1) for the ideal gas $G_{SS}^{ig} = R\cdot T\cdot \ln[P^0 / (R\cdot T)]$, 2) for the unit mole fraction aqueous standard state $G_{SS}^{x} = R\cdot T\cdot \ln[\rho_0 / M_0]$, and 3) for the one mole per kilogram aqueous standard state $G_{SS}^{m} = R\cdot T\cdot \ln[\rho_0 \cdot m_0]$, where M_0 is the molar mass and ρ_0 specific density of water, $P^0 = 0.1$ MPa is the standard pressure and $m^0 = 1\ mol\cdot kg^{-1}$. The determination of functional group contributions for aqueous solute properties used the thermodynamic model proposed by Sedlbauer *et al.* (2000). Conversions between the different possible standard states is ensured by the conversion factor, Λ_P, in Eqn.(2.1). If the unit mole fraction aqueous standard state is applied, Λ_P is

equal to zero for all thermodynamic properties in Eqn.(2.1). If the aqueous standard state is one mole per kilogram solution, Λ_P is zero for all thermodynamic properties except for standard molar Gibbs energy and standard molar entropy, for which

$\Lambda_G = R \cdot T \cdot \ln M_0 \cdot m^0$ (conversion between G_{SS}^x and G_{SS}^m) and

$\Lambda_S = -(\partial \Lambda_G / \partial T)_P = -R \cdot \ln M_0 \cdot m^0$.

The model of Sedlbauer *et al.* (2000) is based on the following equation for standard molar volume,

$$V_{2,i}^0 = d_i \cdot (V_0 - V_{SS}) + \rho_0 \cdot \kappa_0 \cdot R \cdot T \cdot (a_i + b_i \cdot (e^{\vartheta \cdot \rho_0} - 1) + c_i \cdot e^{\theta/T} + \delta_i \cdot (e^{\lambda \cdot \rho_0} - 1)) \quad . \qquad (2.2)$$

Isothermal integration of Eqn.(2.2) from ideal gas standard state to aqueous standard state provides an equation for the hydration Gibbs free energy,

$$\begin{aligned} \Delta_{hyd} G_{2,i}^0 = {} & d_i \cdot (G_0 - G_0^{ig} - \Delta_{hyd} G_{SS}) + \\ & R \cdot T \cdot [(a_i + c_i \cdot e^{\theta/T} - b_i - \delta_i) \cdot \rho_0 + (b_i / \vartheta) \cdot (e^{\vartheta \cdot \rho_0} - 1) + \\ & (\delta_i / \lambda) \cdot (e^{\lambda \cdot \rho_0} - 1)] + H_{2,i}^{corr} - T \cdot S_{2,i}^{corr} \end{aligned} \qquad (2.3)$$

where the last two terms are corrections for the integration through a phase boundary (between ideal gas and liquid). The other standard thermodynamic properties are calculated by simple differentiation,

$$\begin{aligned} \Delta_{hyd} S_{2,i}^0 = -(\partial \Delta_{hyd} G_{2,i}^0 / \partial T)_P = {} & d_i \cdot (S_0 - S_0^{ig} - \Delta_{hyd} S_{SS}) + \\ & R \cdot \theta \cdot c_i \cdot e^{\theta/T} \cdot \tfrac{\rho_0}{T} - R \cdot [(a_i + c_i \cdot e^{\theta/T} - b_i - \delta_i) \cdot \rho_0 + \tfrac{b_i}{\vartheta} \cdot (e^{\vartheta \cdot \rho_0} - 1) \\ & + \tfrac{\delta_i}{\lambda} \cdot (e^{\lambda \cdot \rho_0} - 1)] - R \cdot T \cdot (\tfrac{\partial \rho_0}{\partial T})_P \cdot [a_i + b_i \cdot (e^{\vartheta \cdot \rho_0} - 1) + c_i \cdot e^{\theta/T} + \\ & \delta_i \cdot (e^{\lambda \cdot \rho_0} - 1)] + S_{2,i}^{corr} \end{aligned} \quad , \qquad (2.4)$$

$$\Delta_{hyd}H^0_{2,i} = \Delta_{hyd}G^0_{2,i} + T\cdot\Delta_{hyd}S^0_{2,i} = d_i\cdot(H_0 - H^{ig}_0 - \Delta_{hyd}H_{SS}) + R\cdot T\cdot\theta\cdot c_i\cdot e^{\theta/T}\cdot\frac{\rho_0}{T} - R\cdot T^2\cdot(\frac{\partial p_0}{\partial T})_P\cdot[a_i + b_i\cdot(e^{\vartheta\cdot\rho_0} - 1) + c_i\cdot e^{\theta/T} + \delta_i\cdot(e^{\lambda\cdot\rho_0} - 1)] + H^{corr}_{2,i} \quad (2.5)$$

$$\Delta_{hyd}C^0_{P,2,i} = (\partial H^0_{2,i}/\partial T)_P = d_i\cdot(C_{P,0} - C^{ig}_{P,0} - \Delta_{hyd}C_{P,SS}) - T\cdot(R\cdot c_i\cdot e^{\theta/T}\cdot\theta^2\cdot\frac{\rho_0}{T^3} + 2\cdot R\cdot(\frac{\partial p_0}{\partial T})_P\cdot(a_i + b_i\cdot(e^{\vartheta\cdot\rho_0} - 1) + c_i\cdot e^{\theta/T} + \delta_i\cdot(e^{\lambda\cdot\rho_0} - 1) - c_i\cdot e^{\theta/T}\cdot\frac{\theta}{T}) + R\cdot T\cdot(\frac{\partial p_0}{\partial T})^2_P\cdot(\vartheta\cdot b_i\cdot e^{\vartheta\cdot\rho_0} + \lambda\cdot\delta_i e^{\lambda\cdot\rho_0}) + R\cdot T\cdot(\frac{\partial^2\rho_0}{\partial T^2})_P\cdot[a_i + b_i\cdot(e^{\vartheta\cdot\rho_0} - 1) + c_i\cdot e^{\theta/T} + \delta_i\cdot(e^{\lambda\cdot\rho_0} - 1)]) + C^{corr}_{P,2,i} \quad (2.6)$$

$$\kappa^0_{2,i} = -(\partial V^0_{2,i}/\partial P)_T = d_i\cdot(\frac{M_0}{\rho_0^2}\cdot(\frac{\partial\rho_0}{\partial P})_T - \kappa_{SS}) - R\cdot T\cdot(\frac{\partial\rho_0}{\partial P})^2_T\cdot(\vartheta\cdot b_i e^{\vartheta\cdot\rho_0} + \lambda\cdot\delta_i e^{\lambda\cdot\rho_0}) - R\cdot T\cdot(\frac{\partial^2\rho}{\partial P^2})_T\cdot[a_i + b_i\cdot(e^{\vartheta\cdot\rho_0} - 1) + c_i\cdot e^{\theta/T} + \delta_i\cdot(e^{\lambda\cdot\rho_0} - 1)] \quad (2.7)$$

where ρ_0 is the specific density, κ_0 isothermal compressibility and M_0 molar mass of water, R is the ideal gas constant and T is the temperature in Kelvin. In the equations above, the quantities V_0, G_0, S_0, H_0, and $C_{p,0}$ refer to the molar volume, Gibbs free energy, entropy, enthalpy and heat capacity of pure water. The quantities V^{ig}_0, G^{ig}_0, S^{ig}_0, H^{ig}_0, and $C^{ig}_{P,0}$ refer to the same properties of pure water in the ideal gas state at temperature T and pressure $P^0 = 0.1$ MPa. The coefficients $\vartheta = 0.005\ m^3\cdot kg^{-1}$, $\lambda = -0.01\ m^3\cdot kg^{-1}$ and $\theta = 1500$ K are general and do not depend on the type of functional group. The values of a_i, b_i c_i and d_i are group-specific adjustable parameters and the parameter δ_i is determined by the charge of group i. Values for δ_i have been reported earlier (Sedlbauer *et al.*, 2000),

$$\delta_i = \begin{cases} 0.35 \cdot a_i; & z = 0 \\ 0; & z = 1 \\ -0.645; & z = -1 \end{cases}, \tag{2.8}$$

and are in compliance with the hydrogen convention, $\Xi_2^0[H^+] = 0$. Subtraction of the standard state terms in the first parts of Eqns.(2.2)-(2.7) assures that only one standard state term per free particle in solution is included in the summation of groups in Eqn. (2.1). The thermodynamic standard states are given by,

$$V_{SS} = \kappa_0 \cdot R \cdot T, \tag{2.9}$$

$$\Delta_{hyd} G_{SS} = R \cdot T \cdot \ln[\tfrac{\rho_0 \cdot R \cdot T}{M_0 \cdot P^0}], \tag{2.10}$$

$$\Delta_{hyd} S_{SS} = -R \cdot (\ln[\tfrac{\rho_0 \cdot R \cdot T}{M_0 \cdot P^0}] + 1 - T \cdot \alpha_0), \tag{2.11}$$

$$\Delta_{hyd} H_{SS} = R \cdot T^2 \cdot \alpha_0 - R \cdot T, \tag{2.12}$$

$$\Delta_{hyd} C_{P,SS} = 2 \cdot R \cdot T \cdot \alpha_0 + R \cdot T^2 \cdot (\tfrac{\partial \alpha_0}{\partial T})_P - R, \tag{2.13}$$

$$\kappa_{SS} = -R \cdot T \cdot ((\tfrac{\partial^2 \rho_0}{\partial P^2})_T / \rho_0 - ((\tfrac{\partial \rho_0}{\partial P})_T / \rho_0)^2), \tag{2.14}$$

where $\alpha_0 = -\frac{1}{\rho_0}(\frac{\partial \rho_0}{\partial T})_P$ is the coefficient of thermal expansion of pure water. Correction terms in Eqns. (2.3) – (2.6) are given by,

$$S_{2,i}^{corr} = \begin{cases} e_i \cdot (T - T_c - \frac{T_c^2}{\Theta} \cdot \ln[\frac{T}{T_c}] + \frac{(T_c-\Theta)^2}{\Theta} \ln[\frac{T-\Theta}{T_c-\Theta}]) & z=0; T<T_c \\ C_{S,i} + g_i \cdot (T - T_c) + \ln[\frac{T}{T_c}] \cdot (e_i - g_i \cdot \Theta) \cdot \frac{T_c}{\Theta} + \ln[\frac{T-\Theta}{T_c-\Theta}] \cdot (\Theta - T_c) \cdot \frac{g_i}{\Theta} & z \neq 0; T<T_c \\ 0 & z=0; T \geq T_c \\ C_{S,i} & z \neq 0; T \geq T_c \end{cases} \quad (2.15)$$

$$H_{2,i}^{corr} = \begin{cases} e_i \cdot ((2 \cdot T_c - \Theta) \cdot (T_c - T) + 1/2 \cdot (T^2 - T_c^2) + (T_c - \Theta)^2 \cdot \ln[\frac{T-\Theta}{T_c-\Theta}]) & z=0; T<T_c \\ C_{H,i} + g_i (T^2 - T_c^2)/2 + (T - T_c) \cdot (e_i - g_i \cdot T_c) + e_i \cdot (\Theta - T_c) \cdot \ln[\frac{T-\Theta}{T_c-\Theta}] & z \neq 0; T<T_c \\ 0 & z=0; T \geq T_c \\ C_{H,i} & z \neq 0; T \geq T_c \end{cases} \quad (2.16)$$

$$C_{P,2,i}^{corr} = \begin{cases} e_i \cdot (T - T_c)^2 / (T - \Theta) & z=0; T<T_c \\ (T - T_c) \cdot (e_i / (T - \Theta) + g_i) & z \neq 0; T<T_c \\ 0 & T \geq T_c \end{cases} \quad (2.17)$$

where T_c = 647.126 K is the critical temperature of water and Θ = 228 K is a general constant. The quantities e_i and g_i, for only electrolyte groups, are additional adjustable parameters. The integration constants for the ionic groups, $C_{H,i}$ and $C_{S,i}$, have to be retrieved from experimental hydration enthalpy and entropy data a reference state point (Sedlbauer *et al.*, 2000).

The properties frequently used in calculating chemical equilibria are the partial molar heat capacities and partial molar entropies, $C_{P,2}^0$ and S_2^0, and the Gibbs free energies and enthalpies of formation, $\Delta_f G_2^0$ and $\Delta_f H_2^0$. The relations for these quantities are,

$$C_{P,2}^0 = \Delta_{hyd} C_{P,2}^0 + C_{P,2}^{ig} \quad , \quad (2.18)$$

$$S_2^0 = S_2^0[T_r, P_r] + \int_{T_r}^{T} C_{P,2}^{ig} / T dT + \Delta_{hyd} S_2^0 - \Delta_{hyd} S_2^0[T_r, P_r] \quad , \tag{2.19}$$

$$\Delta_f H_2^0 = \Delta_f H_2^0[T_r, P_r] + \int_{T_r}^{T} C_{P,2}^{ig} dT + \Delta_{hyd} H_2^0 - \Delta_{hyd} H_2^0[T_r, P_r] \quad , \tag{2.20}$$

$$\begin{aligned} &\Delta_f G_2^0 = \Delta_f H_2^0 - T \cdot S_2^0 + T_r \cdot (\sum_{el} S_{el}^0[T_r, P_r] - z/2 \cdot S_{H_2}^0[T_r, P_r]) = \\ &\Delta_f G_2^0[T_r, P_r] - S_2^0[T_r, P_r] \cdot (T - T_r) + \Delta_{hyd} G_2^0 - \Delta_{hyd} H_2^0[T_r, P_r] + \\ &T \cdot \Delta_{hyd} S_2^0[T_r, P_r] + \int_{T_r}^{T} C_{P,2}^{ig} dT - T \cdot \int_{T_r}^{T} C_{P,2}^{ig} / T dT \end{aligned} \quad , \tag{2.21}$$

where $S_2^0[T_r, P_r]$ and $\Delta_f H_2^0[T_r, P_r]$ are experimental values of the partial molar entropy and formation enthalpy in aqueous solution at a reference temperature and pressure of T_r and P_r, and $C_{P,2}^{ig}$ is the solute's ideal gas heat capacity. The expression, $\sum_{el} S_{el}^0[T_r, P_r]$, is the sum of the entropies of the constituting elements in their standard states at T_r and P_r, and $S_{H_2}^0[T_r, P_r]$ is the entropy of hydrogen in its standard state at T_r and P_r. Although it is possible in principle to calculate thermodynamic properties of formation for aqueous nonelectrolyte species without using experimental values of these properties at reference conditions (Sedlbauer *et al.*, 2000), we preferred the formulation in which the values of $S_2^0[T_r, P_r]$ and $\Delta_f H_2^0[T_r, P_r]$ are included, because they can be more easily used for ionic solutes. The integration constants $C_{H,i}$ and $C_{S,i}$ for ionic groups in Eqns.(2.15) and (2.16) do not need to be evaluated. They cancel out in Eqns.(2.19) and (2.20), because they are already included in the experimental values of $S_2^0[T_r, P_r]$ and $\Delta_f H_2^0[T_r, P_r]$.

For convenience the values of $S_{2,i}^0[T_r, P_r]$, $\Delta_f H_{2,i}^0[T_r, P_r]$ and $C_{P,2}^{ig}$ have also been retrieved via group contribution scheme, Eqn.(2.1), with the same groups as those used

for the partial molar thermodynamic properties. For this purpose the experimental results on formation properties reviewed by Amend and Helgeson (1997a), Shock and Helgeson (1990) and Wagman *et al.* (1982) have been used in case of $S^0_{2,i}[T_r, P_r]$ and $\Delta_f H^0_{2,i}[T_r, P_r]$. $C^{ig}_{P,2,}$ group contributions were calculated using the method of Joback (Reid *et al.*, 1987), which gives the ideal gas heat capacity of a compound as a third-order polynomial in absolute temperature,

$$C^{ig}_{P,2} = (-37.93 + \sum_i n_i \cdot \Delta_a) + (0.210 + \sum_i n_i \cdot \Delta_b) \cdot T + (3.91 \cdot 10^{-4} + \sum_i n_i \cdot \Delta_c) \cdot T^2 + (2.06 \cdot 10^{-7} + \sum_i n_i \cdot \Delta_d) \cdot T^3 \quad (2.23)$$

The values of $S^0_{2,i}[T_r, P_r]$, $\Delta_f H^0_{2,i}[T_r, P_r]$, and the coefficients $\Delta_a - \Delta_d$ for our groups are given in Table A1 along with the standard state terms at $T_r = 298.15$ K and $P_r = 0.1$ MPa and the entropies of relevant elements in their standard states at T_r and P_r.

Thermodynamic properties of formation for aqueous amino acids cannot be calculated with parameters from Table A1. The values of $S^0_2[T_r, P_r]$ and $\Delta_f H^0_2[T_r, P_r]$ given by Amend and Helgeson (1997b) are not in good agreement with the functional group additivity assumption, consequently they were not used in calculating our parameters in Table A1. However, Amend and Helgeson (1997b) give a summary of these properties for a set of amino acids, which can be used for each specific compound.

As with any functional group additivity scheme there must be a balance between accuracy and the number of predictable compounds. In order to assure the widest application of the model, only a few compounds (discussed later), were excluded from consideration because of the physical arrangement of their functional groups. With the

exception of amino acids, the functional groups used in this study were limited to the groups contained in the compounds reported by the high temperature measurements of Criss and Wood (1996) and Inglese *et al.* (1996b, 1996c, 1997), which had the effect of allowing each group to be determined over a wide temperature and pressure range. Because of this high temperature and pressure restriction, the following functional groups were used in this study: C (from hydrocarbon groups such as CH_3, CH_2, CH, and C), H (attached to C), COOH, OH, NH_2, $CONH_2$, NH_3^+ and COO^-. For amino acids, the accuracy of representing their properties with NH_3^+ and COO^- functional groups proved to be insufficient and amino acids have been ascribed a special functional group, AMINO (which, in fact, is a combination of NH_3^+ and COO^- groups with quantitatively unknown sterical and electrostatic effects). The determination of the parameters for inorganic ions (Na^+, Cl^-, Br^-), needed for the evaluation of organic electrolytes, has been done previously (Sedlbauer *et al.*, 2000). The determination of parameters for inorganic ions was based on the standard molar volume, heat capacity and isothermal compressibility data for a number of 1-1 electrolytes and was found to yield a good description to 725 K. Thus, these parameters for inorganic ions were not used as additional degrees of freedom in our regression procedure.

The functional group parameters (a_i, b_i, c_i, d_i, e_i for all groups and additional g_i for ionic groups) were regressed using a least-squares procedure to fit Eqn. (2.1) simultaneously to all experimental data as reported in Table 2.1. Experimental errors were not used as weighting factors in the fit. They are not appropriate for this purpose, because compared to the errors in functional group additivity they are considerably lower

Table 2.1: Functional group parameters for aqueous organic solutes.

	$a_i \cdot 10^3$ $m^3 / kg / mol$	$b_i \cdot 10^5$ $m^3 / kg / mol$	$c_i \cdot 10^6$ $m^3 / kg / mol$	d_i	e_i $J / K^2 / mol$	g_i $J / K^2 / mol$
C	-3.6829	-4.8478	-10.206	0.50167	-0.19399	
H	1.9223	6.3555	1.0243	-0.11515	0.01074	
$CONH_2$	-1.4589	-4.5547	-9.1552	2.24596	-0.20309	
$COO^{(-)}$	-656.61	-17.615	6.4012	3.0953	-60.687	0.28352
COOH	-1.3355	1.7165	-14.479	1.4944	-0.32356	
NH_2	3.3633	-8.2574	-5.4818	1.5337	-0.11052	
$NH_3^{(+)}$	-5.2936	7.8583	2.7704	0.17045	-20.413	0.28305
OH	-0.52169	-5.0156	5.7593	1.1057	0.11100	
AMINO	–657.51	11.594	-18.096	1.4227	197.43	-0.61113

at ambient conditions and become comparable at about 500 K for most properties. The weighting scheme employed in the determination of parameters was designed to yield the most reliable predictions over the entire temperature and pressure range, and to balance the impact of different thermodynamic properties on the overall fit. Uncertainties were ascribed to each property in such a way that the value of the average Δ/σ (where Δ is the absolute value of the difference between calculated and experimental point and σ is the prescribed uncertainty of this point) was about unity for each property in the fit. Since the accuracy of experimental data for organic electrolyte species is generally lower than that for nonelectrolyte solutes, the data for electrolytes were considered separately in the weighting scheme. The final estimates of σ for organic compounds were: 1.2% (but at least 0.4 $cm^3 \cdot mol^{-1}$) for V_2^0 of nonelectrolytes, 2.2% (at least 0.7 $cm^3 \cdot mol^{-1}$) for V_2^0 of electrolytes, 4% (at least 5 $J \cdot K^{-1} \cdot mol^{-1}$) for $C_{P,2}^0$ of nonelectrolytes, 7% (at least 9 $J \cdot K^{-1} \cdot mol^{-1}$) for $C_{P,2}^0$ of electrolytes, 10% (at least $4 \cdot 10^{-3}$ $cm^3 \cdot mol^{-1} \cdot MPa^{-1}$) for κ_2^0, 2% (at least 200 $J \cdot mol^{-1}$) for $\Delta_{hyd} H_2^0$, and 1.5% (at least 150 $J \cdot mol^{-1}$) for $\Delta_{hyd} G_2^0$. No experimental results for electrolyte solutes are present in our databases for κ_2^0, $\Delta_{hyd} H_2^0$ and $\Delta_{hyd} G_2^0$. It should be noted that this type of weighting scheme also provides at the same time estimates of the uncertainties in the predictions of these thermodynamic properties for organic compounds, which utilize our group contribution procedure and which are reasonably similar to our training set.

2.4 Accuracy and Performance

In choosing functional groups to represent our results, the most widely applicable groups have been considered at the cost of some accuracy because this choice allows more compounds to be predicted and our data set is too limited to accurately determine a larger number of functional groups. For instance, we used C and H groups instead of the slightly more accurate C, CH, CH_2, and CH_3 groups because the present data set is insufficient to accurately determine the C and CH groups at high temperatures. However contributions of the C and CH groups can be determined to 623 K via a linear combination of the contributions for the C and H groups, even though all the experimental results of compounds containing CH groups are at temperatures below 328 K and a pressure of 0.1 MPa. The advantage of using separate carbon and hydrogen groups far out-weighs the small reduction in the accuracy of the predictions. Preliminary fits were performed using the alternative group division of CH, CH_2, and CH_3. These were found to introduce no significant changes in any of the group contributions with the exception of the CH group (this is expected because the CH group, under that scheme, is determined only from experimental results below 328 K).

In predicting compounds of the form $CH_3(CH_2)_n$-X we used an -X group instead of both a -CH_2X group and a separate group for CH_3X. The use of an -X group instead of a -CH_2X group allows predictions of secondary, tertiary, and quaternary compounds containing -X functional groups (e.g. compounds like 1-butanol, 2-butanol, and 2-isobutanol are all predicted with only the knowledge of C, H, and OH functional groups). We are aware that this simplification leads to somewhat larger uncertainties and as more

data become available, especially for branched compounds, these approximations will not be necessary.

Examination of the $V_{2,i}^0$ and $\Delta_{hyd}C_{P,2,i}^0$ group contributions shows that over the temperature range of 298 to 523 K and pressure range of 0.1 to 100 MPa the standard molar volumes for the polar groups $CONH_2$, COOH, NH_2, and OH show only a small temperature and pressure dependence (about 5 $cm^3 \cdot mol^{-1}$ being the largest difference). Given the slight temperature and pressure dependencies of these groups, assuming $V_{2,i}^0$ and $\Delta_{hyd}C_{P,2,i}^0$ to be constant between 298 K and 523 K for polar functional groups not included in this study (due to the lack of experimental data) would be appropriate and is in agreement with the findings of Criss and Wood (1996). Above 523 K, $V_{2,i}^0$ and $\Delta_{hyd}C_{P,2,i}^0$ for these polar groups become increasingly negative as the critical point is approached and appear to be a regular family of curves as expected (Hnedkovsky *et al.*, 1996) (Figure 2.1). Group contributions for the NH_3^+, COO^- and AMINO groups show the large negative slopes at higher temperatures for $V_{2,i}^0$ and $\Delta_{hyd}C_{P,2,i}^0$ which are characteristic of ionic species (Figure 2.2). The amino acid functional group is far from being a mere superposition of the NH_3^+, COO^- groups especially at lower and medium temperatures, where it resembles more the behavior of polar groups like COOH. As expected, our definition of amino acid group is more positive than the sum of the separated ions because the ion pair has lower electrostriction than the ions. The difference increases with temperature because the effects of electrostriction increase with decreasing solvent dielectric constant. However, the AMINO functional group should be

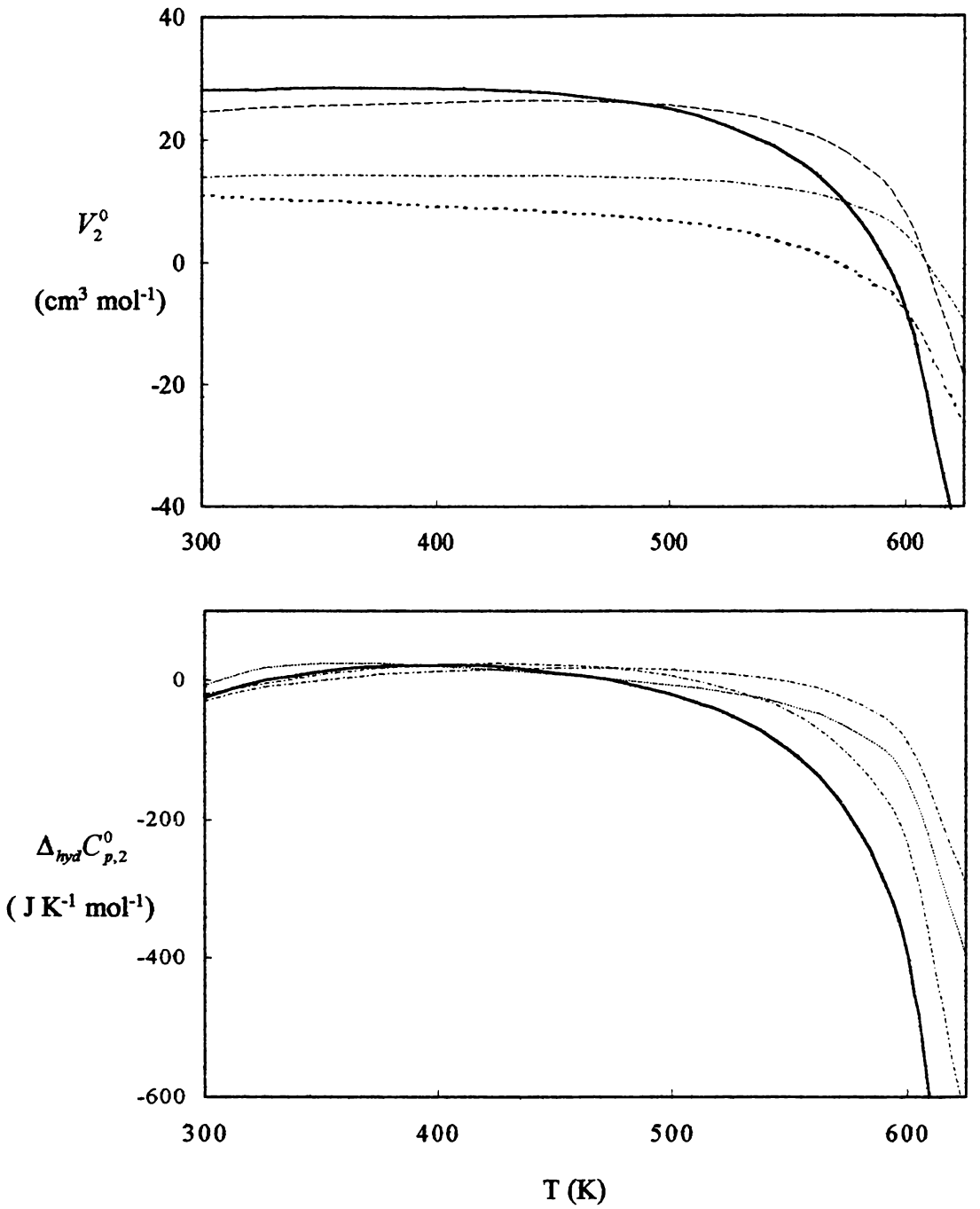

Figure 2.1: Functional group contributions to the partial molar volumes and hydration heat capacities at infinite dilution for $CONH_2$, COOH, NH_2 and OH as a function of temperature. ____ - $CONH_2$; _ _ _ - COOH ; __ _ __ - NH_2; ..... - OH.

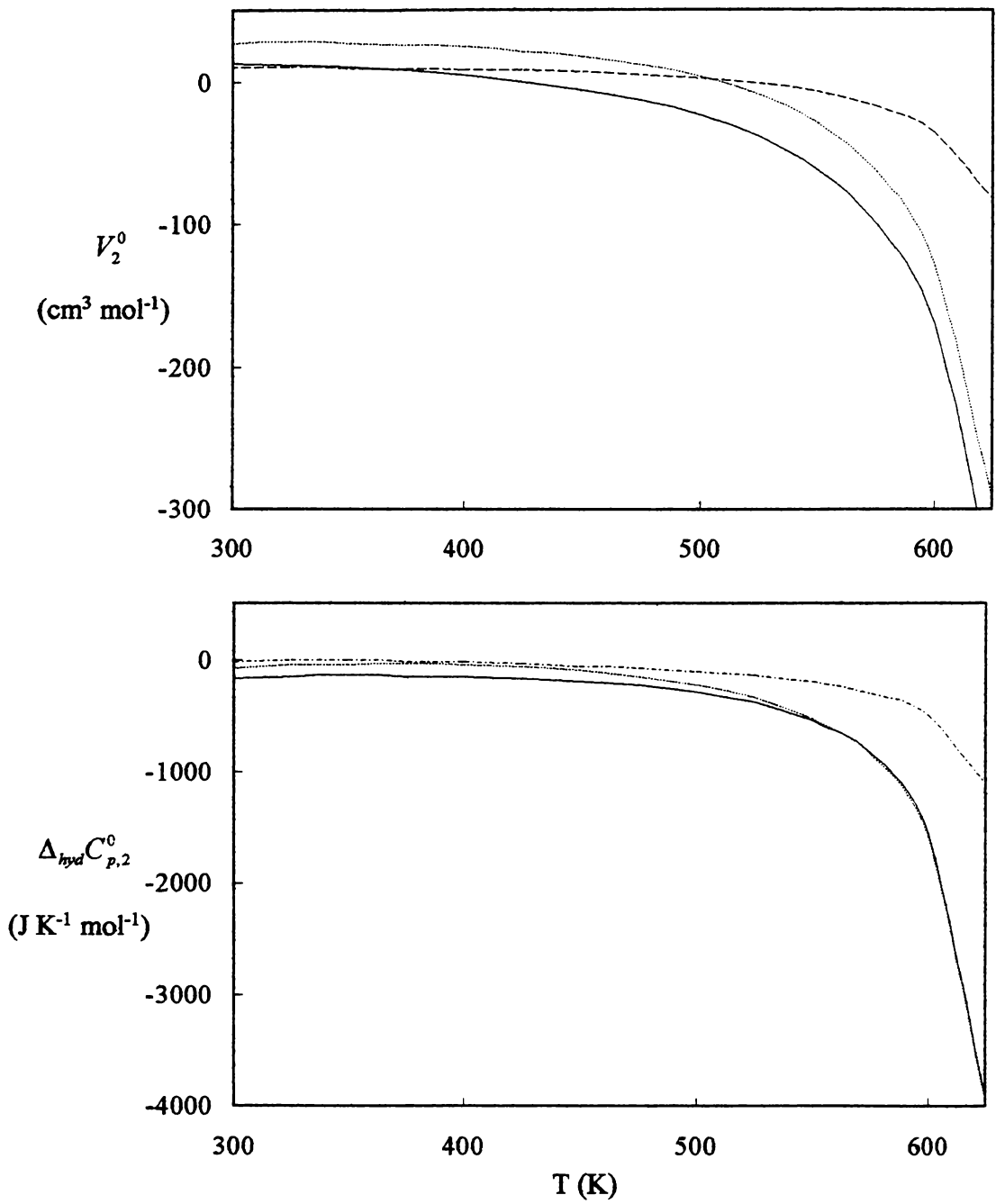

Figure 2.2: Functional group contributions to the partial molar volumes and hydration heat capacities at infinite dilution for COO^-, NH_3^+ and amino acid functional group as a function of temperature. ____ - COO^-; _ _ _ - NH_3^+; - - - - - amino acid group.

used with great caution because the distance between the ions and the strength of any electrostriction effects will depend on the other functional groups in the amino acid. In addition, the force between the ions also changes with dielectric constant so the average distance apart may be temperature and pressure dependant.

The group contributions for $V_{2,i}^0$ and $\Delta_{hyd}C_{P,2,i}^0$ of the C, CH, CH_2, and CH_3 groups to 623 K by linear combination of the C and H groups are given in Figure 2.3. While the temperature dependence of CH and CH_2 groups is small up to 500 K, the C and CH_3 groups exhibit larger changes. The contributions for the CH_2 and CH_3 groups become increasingly positive as the critical point is approached as one might expect for non-polar species.

It is interesting to examine the plot of $\Delta_{hyd}G_2^0(T)_P$ for a series of alkanes (Figure 2.4). In physical terms, $\Delta_{hyd}G_2^0(T)_P$ is a measure of how much free energy is required to transfer a solute molecule from the ideal gas standard state to an aqueous solution. In Figure 2.4, it can be seen that initially $\Delta_{hyd}G_2^0(T)_P$ increases with increasing temperature for all alkanes shown, signaling that it is more difficult to transfer an alkane molecule into water at these conditions. Also the magnitude of $\Delta_{hyd}G_2^0(T)_P$ is growing proportionally to the chain length as expected. The maximum in $\Delta_{hyd}G_2^0(T)_P$ that is reached between ~400 K – 600 K indicates that the relative strength of hydrophobic solvation reaches its maximum value in this region and implies that the strength of the so-called hydrophobic effects reaches its peak at moderate to high temperatures.

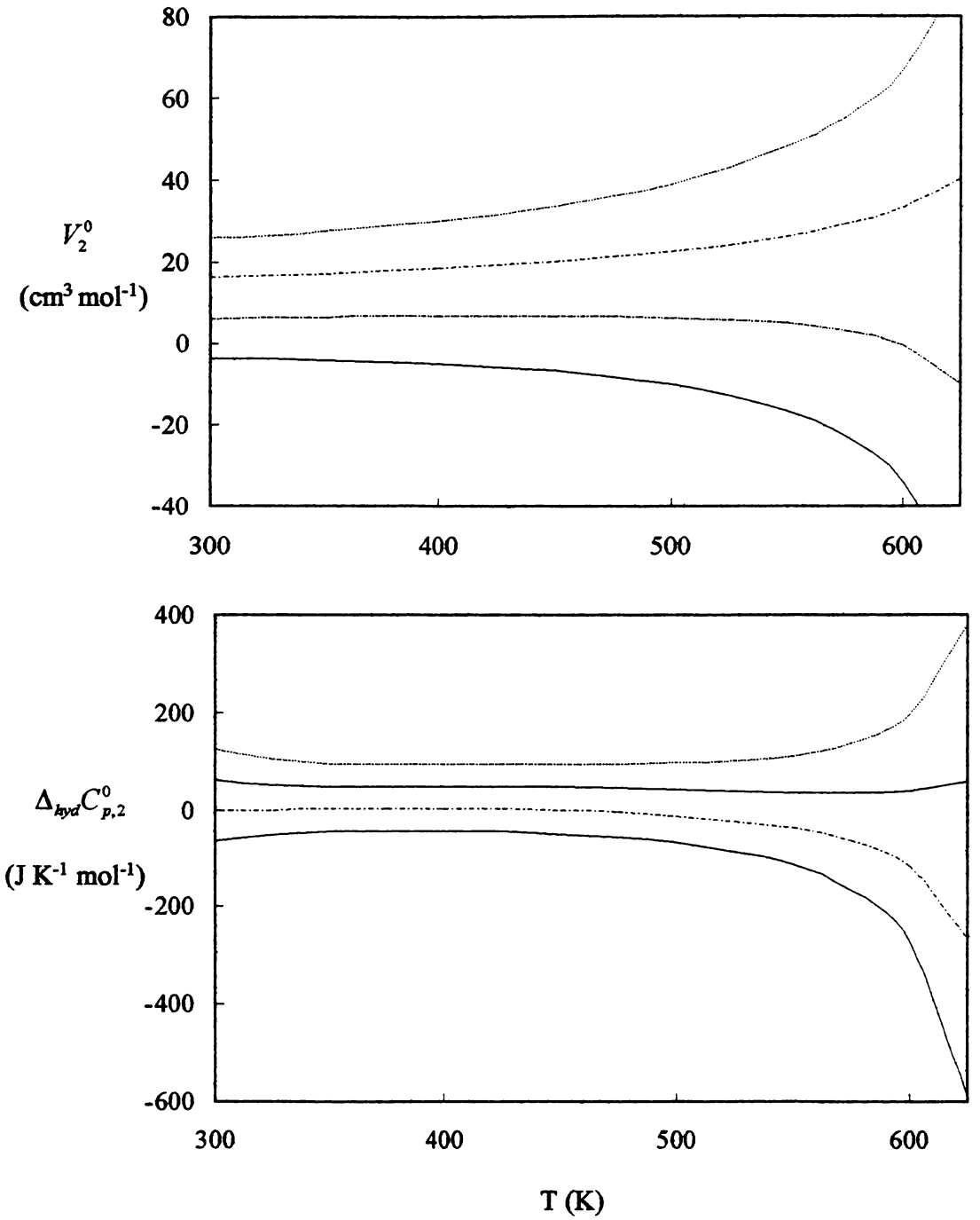

Figure 2.3: Functional group contributions to the partial molar volumes and hydration heat capacities at infinite dilution for C, CH, CH_2 and CH_3 as a function of temperature. ____ - C; _ _ _ - CH; ___ _ ___ - CH_2; ..... - CH_3.

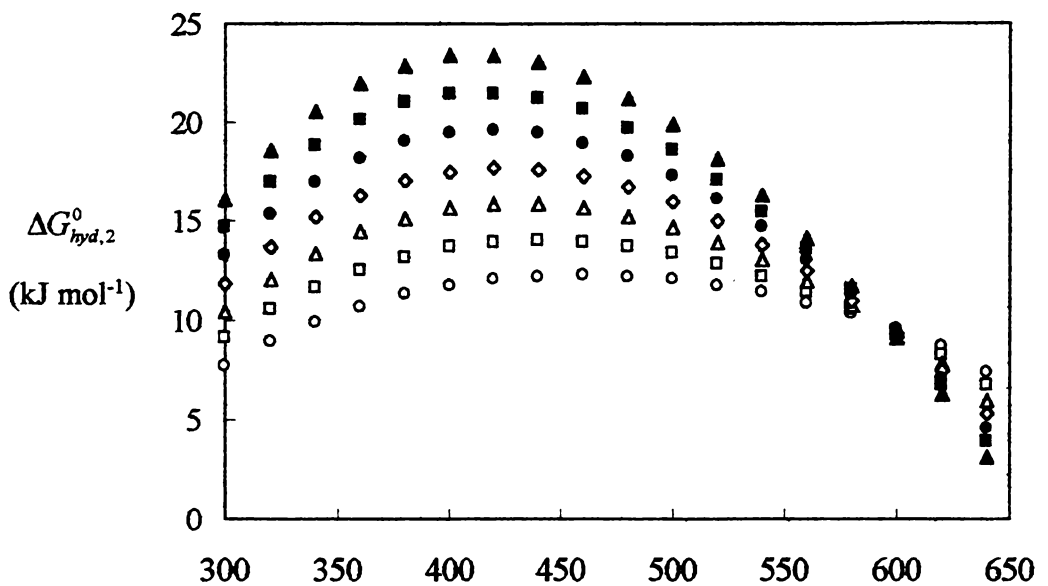

Figure 2.4: Predicted Gibbs energies of hydration for a series of alkanes along the P=28 MPa isobar.O – Methane; □ – Ethane; Δ – Propane; ◇ – Butane; ● – Pentane; ■ – Hexane; ◆- Heptane.

The cause of this maximum from an analytical standpoint can be traced back to the group contributions of CH_3 and CH_2, which both exhibit a maximum in $\Delta_{hyd}G_2^0(T)_P$. Note that for propane, ethane and methane only a weak or no maximum occurs. This is because the standard state term is dominating in these systems. However, as the number of CH_2 monomer units is increased, the temperature dependence of the CH_2 group becomes significant, creating the observed trends.

At higher temperatures $\Delta_{hyd}G_2^0(T)_P$ begins to decrease, implying an increase in the solubility, which is most likely due to the decreasing dielectric constant and hydrogen-bonding structure of the surrounding water solvent. At high enough temperatures a crossover for the alkane curves is observed. The intriguing result of this prediction is that as the critical temperature is approached it is easier to solvate a longer alkane chain than a smaller chain. These predictions will be examined in greater detail in part 4. It should be pointed out, however, that our functional group scheme was adjusted only for molecules with carbon chain length under 12 and therefore one must be very careful when extending the previous statement to larger molecules.

The validity of the extrapolations for the standard thermodynamic properties to elevated conditions and the assumption of group additivity can be examined by the predictions of the standard molar volumes and standard molar heat capacities for methane, via the contributions of the C and H groups. These predictions (Figure 2.5) are in good agreement with the experimental results (Hnedkovsky *et al.*, 1996, 1997), although the results of Hnedkovsky *et al.* were not included in the regression procedure.

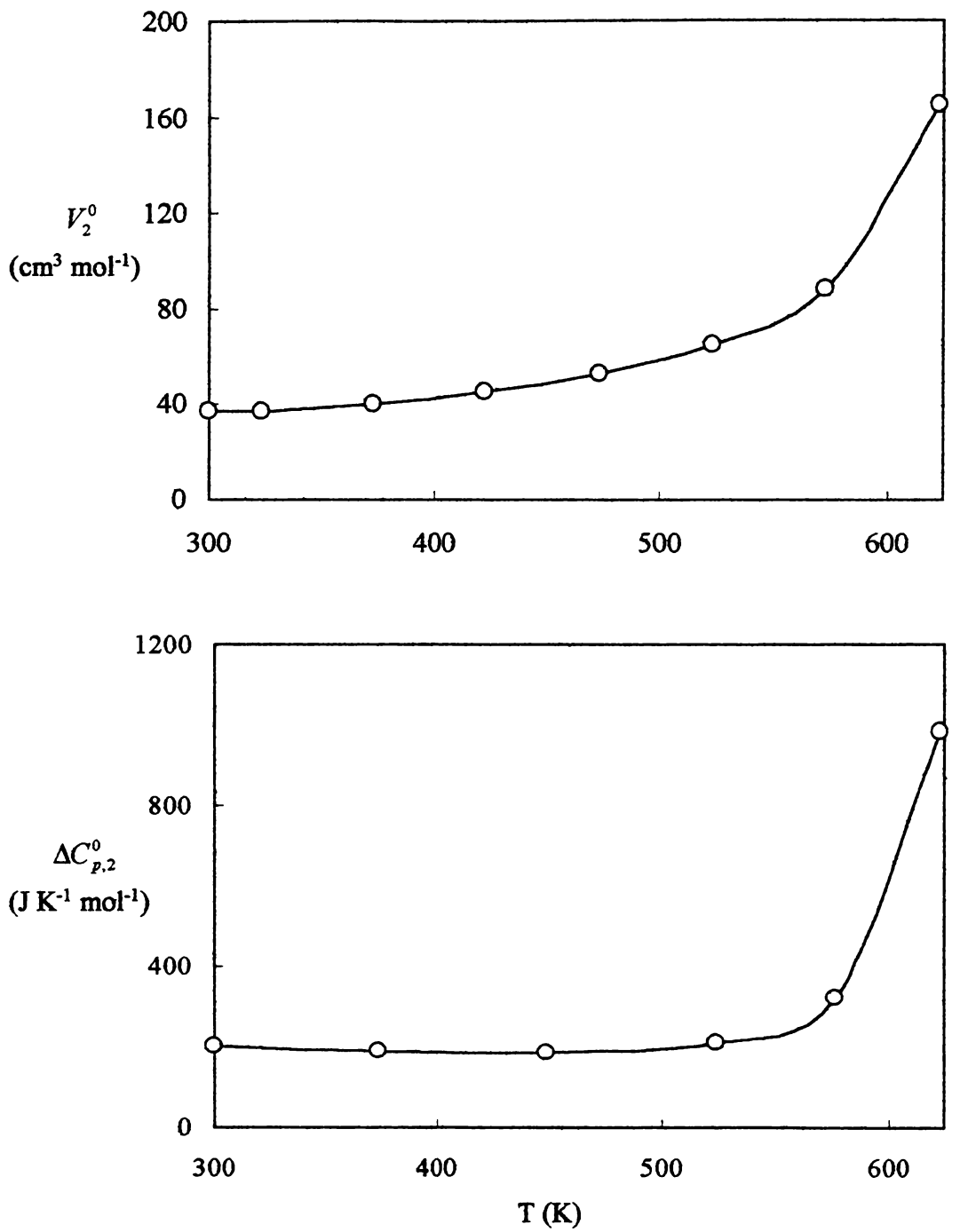

Figure 2.5: Predictions of partial molar volumes and partial molar heat capacities methane at infinite dilution. O - experimental results of Hnedkovsky et al. (1996; 1997); _____ - Eqn.(2.1).

This also provides us with increased confidence that the predictions of the CH group from the C and H groups are reasonably accurate. Another test of extrapolating ability is comparison of the predictions for glycine with new experimental data (not included in the regression) reported by Hakin *et al.* (1998) to 472 K and 30 MPa. The comparison is presented in Table 2.2. The predicted values of partial molar volumes are not too far from experimental results, but the important difference between the two data sets is that experimental data display increase of V_2^0 with temperature and a decrease with pressure, which is a typical behavior for nonelectrolyte solutes, while predictions suggest opposite behavior. The trend in predictions can be explained by the ionic nature of amino acid functional group as it shown in Figure 2.2. It should be noted that data at lower temperatures are in agreement with the predictions (*e.g.*, Chalikian *et al.* (1994) reported electrolyte-like increase in volume with pressure for glycine and alanine). When included in a test regression, the data of Hakin *et al.* could not be described together with that for the low temperature data with better accuracy than that achieved by prediction as presented in Table 2.2, indicating some disagreement between the low and high temperature data. As discussed above the amino group should be used with caution because the electrostriction may depend on neighboring functional groups, temperature, and pressure. In fact, this may be the reason for our inability to fit Hakin's results. The accuracy of predictions for amino acids will remain unclear until some other high temperature results become available and either confirm or disprove the trends presented for glycine by Hakin *et al.*

The only other work comparable to this study was performed by Amend and Helgeson (1997). Their predictions are not expected to be as accurate as the present

Table 2.2: Partial molar volumes at elevated conditions of glycine (Hakin *et al.*, 1998) compared with the predictions from Eqn.(2.1).

T / K	P / MPa	$V^0_{2,Exp}$ [a] / $cm^3 \cdot mol^{-1}$	$V^0_{2,Calc}$ [b] / $cm^3 \cdot mol^{-1}$
397.75	10.0	43.25	44.24
397.00	20.0	42.15	44.58
398.00	30.0	36.49	44.85
423.02	10.0	44.00	43.18
423.47	20.0	41.31	43.59
422.94	30.0	38.06	44.01
472.93	10.0	44.28	37.95
472.85	20.0	43.66	39.00
472.03	30.0	39.81	40.00

[a] experimental values of Hakin *et al.* (1998)
[b] Eqn.(2.1)

predictions at temperatures over 373 K, because their study was submitted before the high temperature experimental studies of Criss and Wood (1996), Inglese *et al.* (1996a, 1996b, 1996c, 1997) were published. The equation of state parameters of the revised HKF model given by Amend and Helgeson (1997) were used to compare their predictions with those of Eqn.(2.1) and with the standard molar volume results reported by Criss and Wood (1996) and with the standard molar heat capacity results for the same compounds (Inglese *et al.* 1996b, 1996c, 1997) (Table 2.3). At high temperatures, the extrapolations of Amend and Helgeson differ significantly from the experimental results in most cases, but their predictions are still reasonable given the limited data set on which their calculations were based. Predictions of Eqn.(2.1) are generally very close to experimental values for standard molar volumes. For standard molar heat capacities the present predictions are, on average, only marginally better than the predictions of Amend and Helgeson. The reason for the lower accuracy of heat capacity predictions from Eqn.(2.1) is clearly in the origin of the model for heat capacity (Sedlbauer *et al.*, 2000). Calorimetric properties are described in this model by one adjustable parameter (two for the ions) additional to parameters of volumetric model, which is insufficient for high accuracy in the heat capacity correlation and prediction. On the other hand, the thermodynamic property of practical interest, the chemical potential, is quite insensitive to some changes in the properties at the derivative level. This can be seen clearly on Figure 2.6, which depicts a comparison between the predictions by Amend and Helgeson and from Eqn.(2.1) for formation Gibbs energy (chemical potential) of two randomly chosen solutes, ethane and butanol. Regardless of the differences in predicted standard molar volumes and standard molar heat capacities for these compounds from the two

Table 2.3: Differences of the experimental values (28 MPa) of the partial molar volumes (Criss and Wood, 1996) and partial molar heat capacities (Inglese *et al.*, 1996b,c; 1997) from calculations by Eqn.(2.1) and by Amend and Helgeson (1997a).

T	$\Delta V_2^{0\,a}$	$\Delta V_2^{0\,b}$	$\Delta C_{p,2}^{0\,c}$	$\Delta C_{p,2}^{0\,d}$
°C	$cm^3 \cdot mol^{-1}$	$cm^3 \cdot mol^{-1}$	$J \cdot K^{-1} \cdot mol^{-1}$	$J \cdot K^{-1} \cdot mol^{-1}$
		Propanol		
25	0.5	1.3		
30			-2	-41
100	-0.5	2.2	12	-4
175	-0.5	9.2	5	-12
250	1.4	20.9	48	-25
		1,4-Butanediol		
25	0.6	-0.1		
30			-6	26
100	-0.7	3.5	0	39
175	-1.6	8.1	-1	38
250	0.9	17.6	50	31
		1,6-Hexanediol		
25	0.0	-0.5		
30			-13	6
100	-1.1	3.2	9	50
175	-2.3	11.8	-23	16
250	-0.8	24.0	54	-4
		Propylamine		
25	1.0	2.1		
30			25	-7
100	-0.2	5.1	22	4
175	-0.6	13.4	-6	3
250	0.4	27.3	6	6
		1,4-Butanediamine		
25	2.5	2.3		
30			-34	23
100	-0.1	8.4	-7	29
175	1.4	16.8	-34	68
250	3.9	35.5	24	152

Table 2.3. (Continued)

T	$\Delta V_2^{0\,a}$	$\Delta V_2^{0\,b}$	$\Delta C_{p,2}^{0\,c}$	$\Delta C_{p,2}^{0\,d}$
°C	$cm^3 \cdot mol^{-1}$	$cm^3 \cdot mol^{-1}$	$J \cdot K^{-1} \cdot mol^{-1}$	$J \cdot K^{-1} \cdot mol^{-1}$
		1,6-Hexanediamine		
25	0.8	0.9		
30			-2	41
100	-0.5	9.0	-6	31
175	-2.4	20.2	-65	37
250	-4.2	35.5	0	88
		Propionamide		
25	0.4	0.5		
30			-2	-20
100	-0.1	0.9	-10	-10
175	-0.3	4.7	-14	-8
250	0.0	7.3	7	-63
		Acetic acid		
30			2	-29
100			11	-1
175			-6	-5
250			3	-35
		Propanoic acid		
25	0.1	0.7		
30			3	-35
100	0.5	1.1	4	-8
175	0.1	5.9	-33	-32
250	-0.4	11.8	7	-51
		Succinic acid		
25	0.5	-0.7		
30			24	68
100	0.7	-0.5	6	56
175	0.6	2.4	-19	65
250	2.0	3.0	11	22

Table 2.3. (Continued)

T / °C	$\Delta V_2^{0\,a}$ / $cm^3 \cdot mol^{-1}$	$\Delta V_2^{0\,b}$ / $cm^3 \cdot mol^{-1}$	$\Delta C_{p,2}^{0\,c}$ / $J \cdot K^{-1} \cdot mol^{-1}$	$\Delta C_{p,2}^{0\,d}$ / $J \cdot K^{-1} \cdot mol^{-1}$
		Adipic acid		
25	0.7	-0.2		
100	1.0	1.0		
175	-0.4	6.0		
250	-3.4	6.7		
		Propylamine hydrochloride		
25	0.7			
30			-2	
100	0.3		12	
175	-0.1		-6	
250	0.0		1	
		Sodium acetate		
25	0.6			
30			1	
100	0.5		-2	
175	-0.2		7	
250	-0.8		-10	
		Sodium propanoate		
25	-1.0			
30			14	
100	-1.6		5	
175	-2.5		-22	
250	-2.3		-1	

[a] $\Delta V_2^0 = V_{2,Exp}^0 - V_{2,Calc}^0$ (Eqn.(2.1))

[b] $\Delta V_2^0 = V_{2,Exp}^0 - V_{2,Calc}^0$ (Amend and Helgeson, 1997a)

[c] $\Delta C_{p,2}^0 = C_{p,2,Exp}^0 - C_{p,2,Calc}^0$ (Eqn.(2.1))

[d] $\Delta C_{p,2}^0 = C_{p,2,Exp}^0 - C_{p,2,Calc}^0$ (Amend and Helgeson, 1997a)

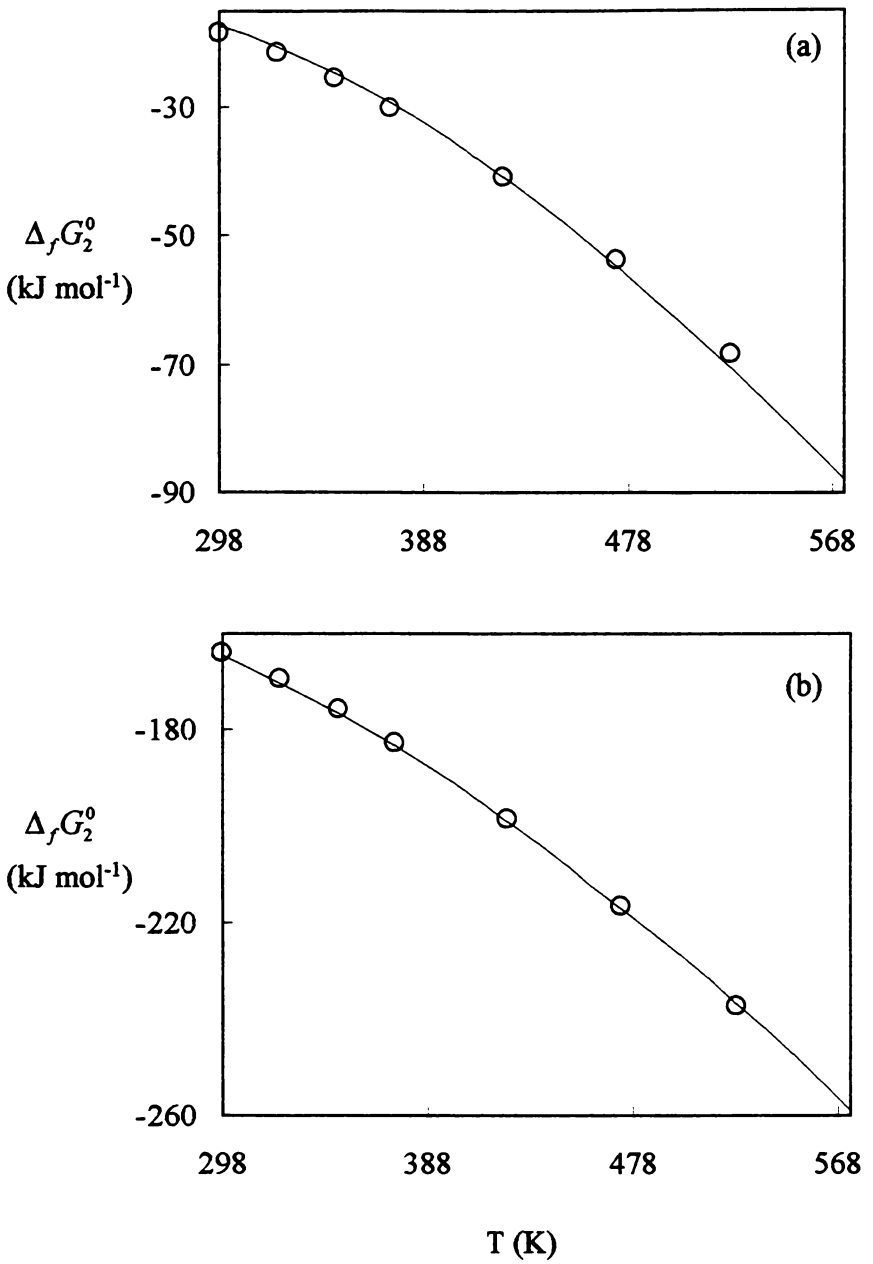

Figure 2.6: Predictions of formation Gibbs energies for ethane and butanol at infinite dilution. a) ethane, b) butanol, _____ - Eqn.(2.1); O - Amend and Helgeson (1997 a).

models, the chemical potentials are almost identical. Predictions of equilibrium constant for dissociation reactions of aliphatic carboxylic acids are shown in Figure 2.7 for the examples of acetic and propanoic acids (comparison with Amend and Helgeson was not possible, since their predictions did not include organic ions; the same applies to comparisons with salts in Table 2.3). Above 550 K we expect the predictions of the present equation for various standard thermodynamic properties to be more accurate than the revised HKF equation even when both equations are parametrized using the same experimental data set because Eqn.(2.1) is more accurate in fitting experimental data near the critical point and is consistent with critical scaling theory near the critical point (Sedlbauer *et al.*, 2000).

The dominant source of error for all properties in this study (perhaps with the exception of the partial molar compressibility) is the assumption of functional group additivity. The inaccuracies in functional group additivity near 298 K are well documented (Cabani *et al.*, 1981; Hoiland, 1986). The assumption of functional group additivity is not as accurate as when near neighbor interactions are strong, *i.e.* when polar, charged, or terminal groups are in close proximity to each other. Our estimated error for all properties due to functional group additivity for compounds with only one polar group and at least 2 carbon atoms (*e.g.*, propanol, propanoic acid and propanamine) is about 50% lower than the uncertainty limits stated in section 2.3. The estimated error is a little larger than the limits in section 2.3 for compounds with two polar groups such as hexanediol and hexanediamine. In extreme cases, compounds with two or more polar groups close together (such as tartaric acid) may produce errors two and perhaps four times larger than the stated limits. About ten experimental results for compounds of this

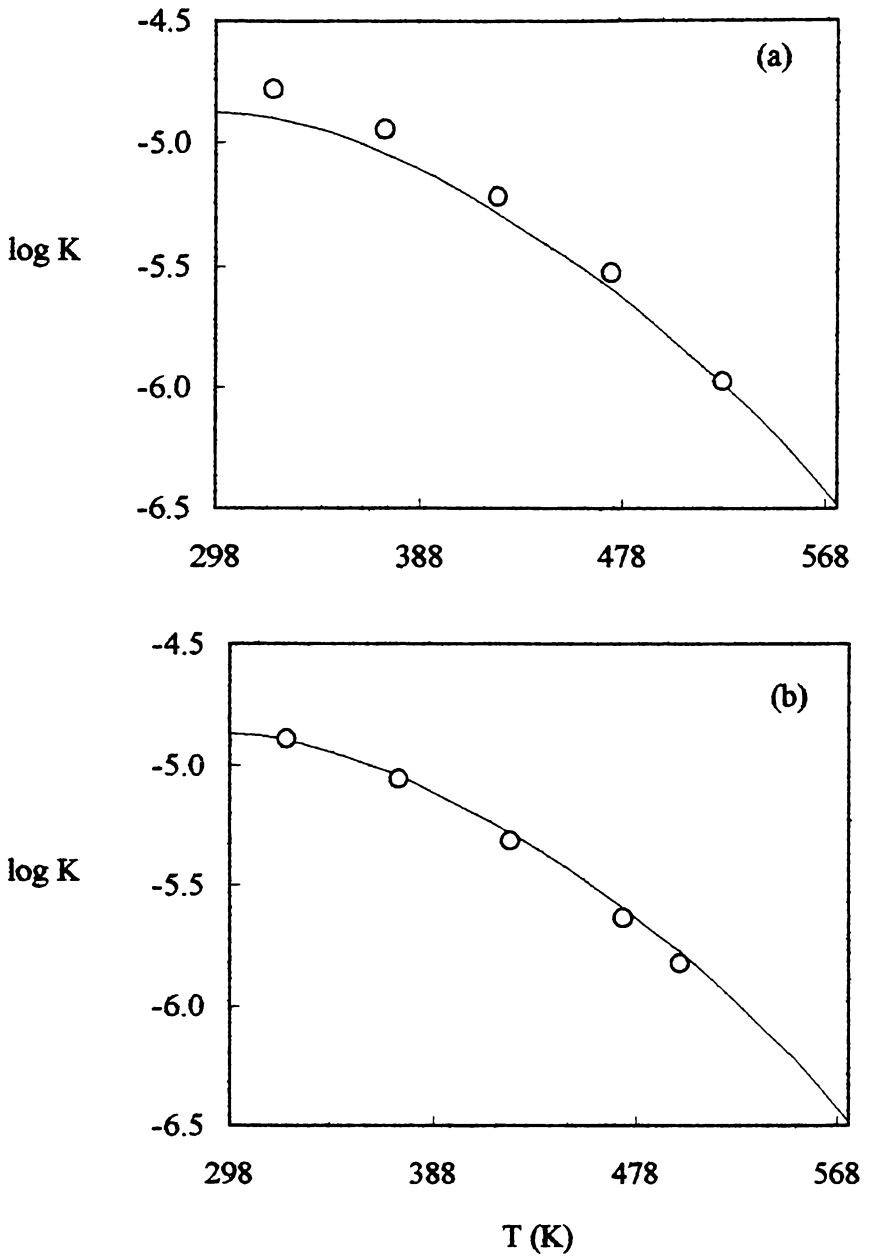

Figure 2.7: Predictions of of dissociation reactions for aliphatic carboxylic acids. a) acetic acid, b) propanoic acid, O - experimental results of Ellis (1963); _____ - Eqn.(2.1).

type (tartaric acid, aspartic acid, serine, urea, ethanediol) were therefore excluded from our regression calculations, because they would unnecessarily affect the accuracy of the functional group contributions. Another limitation of the group additivity assumption is that it predicts the same thermodynamic properties for isomers containing identical functional groups. For example, in the functional group scheme used in this study both leucine and iso-leucine have the same functional group composition, but the partial molar volumes of the two compounds differ experimentally by about 2 $cm^3 \cdot mol^{-1}$ at standard conditions. For partial molar compressibilities not only group additivity assumption, but also the experimental errors and model errors contribute to the large uncertainty limits, especially at low temperatures. Estimating uncertainties in predictions above 520 K where we have no experimental results is difficult. Examination of our predictions and the findings of Sedlbauer *et al.* (2000) allow us to give a rough estimate of the increase in uncertainties when the data are extrapolated. The uncertainties are expected to increase by about a factor of two as temperature increases from 520 K to 620 K.

While the discussion of expected uncertainties regarded derivative thermodynamic properties, it should be noted that when calculating partial molar entropies and enthalpies and Gibbs energies of formation, additional errors in the values of these properties at reference conditions adds to the overall uncertainty. When using our functional group scheme at T_r and P_r with parameters from Table 2.2, the average error in calculating $S_2^0[T_r, P_r]$ was about 2 $J \cdot K^{-1} \cdot mol^{-1}$ (7 $J \cdot K^{-1} \cdot mol^{-1}$ maximum) and for $\Delta_f H_2^0[T_r, P_r]$ the average error was 1 $kJ \cdot mol^{-1}$ (5 $kJ \cdot mol^{-1}$ maximum) (these errors

apply for organic solutes from which the functional groups were determined, as discussed in Appendix A).

2.5 Conclusion

In this part, an equation of state based on a simple functional group additivity scheme for standard molar thermodynamic properties has been presented and been shown to provide a good description of the available experimental data. The choice of functional groups was affected by the set of available high temperature experimental results and reflects a need of predictions for a wide variety of organic compounds. Eqn. (2.1) should allow predictions of chemical potentials at infinite dilution of aqueous organic solutes to 620 K and 100 MPa without large a decrease in accuracy. The databases of experimental data have been collected and are being presented.

The predicted behavior of the free energy of hydration for the alkanes is non-intuitive and the reversal in the isobaric free energy of hydration trends suggests that the traditional picture of solvation (*i.e.* the creation of a cavity volume) is not complete at high temperatures. However it is not clear from the available experimental data whether these predictions represent a physically meaningful effect or are simply an artifact of the least-squares regression. In an attempt to ascertain the validly of these predictions (without performing an expensive and difficult experiment) we have chosen to employ molecular dynamic simulation; the results of which will be discussed in the forthcoming parts.

References

Ackermann, von T. and Schreiner, F. 1958. *Zeitschrift fur Elektrochemie*, **62**, 1143.

Alexander, D.M. 1959. *J.Chem.Eng.Data*, **4**, 252.

Allred, G.C. and Wooley, E.M., 1981. *J.Chem.Thermodyn.*, **13**, 155.

Amend, J. P. and Helgeson, H. C.1997a. *Geochim.Cosmochim.Acta*, **61**, 11.

Amend, J. P. and Helgeson, H. C. 1997b. *J. Chem. Soc, Faraday Trans.*, **93**, 1927.

Ben-Naim, A., 1972. *Hydrophobic Interactions*, Plenum Press, New York.

Ben-Naim, A., 1987. *Solvation Thermodynamics*, Plenum Press, New York.

Cabani, S.; Gianni P.; Mollica, V. and Lepori, L. 1981. *J.Sol.Chem.*, **10**, 563 .

Chalikian, T. V.; Sarvazyan, A.P. and Breslauer, K.J. 1993. *J.Phys.Chem.*, **97**, 13017.

Chalikian, T.V.; Sarvazyan, A.P.; Funck, T.; Cain, C.A. and Breslauer, K.J. 1994. *J.Phys.Chem.*, **98**, 321.

Criss, C.M. and Wood, R.H. 1996. *J. Chem. Thermodyn.*, **28**, 723.

Dec, S.F. and Gill, J.S. 1983. *J.Sol.Chem.*, **131**, 27.

DiPaola, G. and Belleau, B. 1977. *Can.J.Chem.*, **55**, 3825.

DiPaola, G. and Belleau, B. 1978. *Can.J.Chem.*, **56**,1827.

Duke, M. M.; Hakin, A.W.; McKay, R.M. and Preuss, K.E. 1994. *Can.J.Chem.*, **72**, 1489.

Friedman, M. E. and Scheraga, H.A. 1965. *J.Phys.Chem.*, **69**, 3795.

Gianni, P. and Lepori, L. 1996. *J.Sol.Chem.*, **25**, 1.

Hakin, A. W.; Duke, M.M.; Marty, J.L, and Preuss, K.E. 1994a. *J.Chem.Soc.Faraday Trans.*, **90**, 2027.

Hakin, A.W.; Duke, M.M.; Klassen, S.A.; McKay, R.M. and Preuss, K.E. 1994b.

Can.J.Chem., **72**, 362.

Hakin, A. W.; Duke, M.M.; Groft, L.L.; Marty, J.L. and Rushfeldt., M.L. 1995. *Can.J.Chem.*, **73**, 725.

Hakin, A.W.; Copeland, A.K.; Liu, J.L.; Marriot, R.A. and Preuss, K.E. 1997. *J.Chem.Eng. Data*, **42**, 84.

Hakin, A.W.; Daisley, D.C.; Delgado, L.; Liu, J.L.; Marriott, R.A.; Marty, J.L. and Tompkins, G. 1998. *J.Chem.Thermodyn.*, **30**, 583.

Hill, P.G. 1990. *J.Phys.Chem.Ref.Data*, **19**, 1233.

Hnedkovsky, L.; Wood, R.H. and Majer, V., 1996. *J.Chem.Thermodyn.*, **28**, 125.

Hnedkovsky, L. and Wood, R.H. 1997. *J.Chem.Thermodyn.*, **29**, 731.

Hoiland, H. and Vikingstad, E. 1975. *J.Chem.Soc.Faraday Trans.*, **71**, 2007.

Hoiland, H. and Vikingstad, E. 1976. *Acta Chem.Scand.*, **30**, 182.

Hoiland, H. 1980. *J.Sol.Chem.*, **9**, 857.

Hoiland, H. 1986. *Thermodynamic Data for Biochemistry and Biotechnology*.

Hung, H. and Verrall, R.E. 1994. *J.Sol.Chem.*, **23**, 925.

Inglese, A.; Robert, P.; De Lisi, R. and Milioto, S., 1996a. *J.Chem.Thermodyn.*, **28**, 873.

Inglese, A. and Wood, R.H. 1996b. *J.Chem.Thermodyn.*, **28**, 1059.

Inglese, A.; Sedlbauer, J. and Wood, R.H., 1996c. *J.Sol.Chem.*, **25**, 849.

Inglese, A.; Sedlbauer, J.; Yezdimer, E. M. and Wood, R.H. 1997. *J.Chem.Thermodyn.*, **29**, 517.

Jolicoeur, C. and Lacroix,G. 1976. *Can.J.Chem.*, **54**, 624.

Kaulgud, M.V.; Bhagde, V.S. and Shrivastava, A. 1982. *J.Chem.Soc.Faraday Trans*, **78**, 313.

Kawaizumi, F.; Noguchi, T. and Miyahara, Y. 1977. *Bull.Chem.Soc.Jpn.*, **50**, 1687.

Kikuchi, M.; Sakura, M. and Nitta, K. 1995. *J.Chem.Eng.Data*, **40**, 935.

Kiyohara, O.; Perron, G. and Desnoyers, J.E. 1975. *Can.J.Chem.*, **35**, 3263.

King, E. J. 1969. *J.Phys.Chem.*, **73**, 1220.

Lepori, L. and Mollica, V. 1980. *J.Phys.Chem.*, **123**, 51.

Maham, Y.; Teng, T.T.; Helper, L.G. and Mather, A.E. 1994. *J.Sol.Chem.*, **23**, 195.

Makhatadze, G.I. and Privalov,P.L. 1989. *J.Sol.Chem.*, **18**, 927.

Makhatadze,G.I.; Medvedkin, V.N. and Privalov, P.L. 1990. *Biopolymers*, **30**, 1001.

Makhatadze, G.I. and Privalov, P.L. 1990. *J.Mol.Biol.*, **213**, 375.

Millero, F. J.; Lo Surdo, A. and Shin, C. 1978. *J.Phys.Chem.*, **82**, 784.

Mishra, A. K. and Ahluwalia, J.C. 1984. *J.Phys.Chem.*, **88**, 86.

Mizuguchi, M.; Sakurai, M. and Nitta, K. 1997. *J.Sol.Chem.*, **26**, 579.

Naghibi, H.; Dec, S.F. and Gill, S.J. 1987. *J.Phys.Chem.*, **91**, 254.

Nakajima, T.; Komatsu, T. and Nakagava, T. 1975. *Bull.Chem.Soc.Jpn.*, **48**, 783.

Nichols, N.; Skold, R.; Spink C. and Wadso, I. 1976. *J.Chem.Thermodyn.*, **8**, 993.

O'Connell, J. P.; Sharygin, A.V. and Wood, R.H. 1996. *Ind.Eng.Chem.Res.*, **35**, 2808.

Reid, R.C.; Prausnitz, J.M. and Poling, B.E. 1987. *The properties of gases and liquids.* McGraw-Hill Book Company, New York.

Rosenholm, J. B. and Hepler, L.G. 1984. *Thermochimica Acta*, **81**, 381.

Roux, G.; Roberts, D.; Perron, G. and Desnoyers, J.E. 1980. *J.Sol.Chem.*, **9**, 629.

Sakurai, M. 1973. *Bull.Chem.Soc.Jpn.*, **46**, 1596.

Sakurai, M.; Komatsu, T. and Nakagawa, T. 1975a. *Bull.Chem.Soc.Jpn.*, **48**, 3491.

Sakurai, M.; Komatsu, T. and Nakagawa, T. 1975b. *J.Sol.Chem.*, **4**, 511.

Sakurai, M. 1987. *Bull.Chem.Soc.Jpn.*, **60**, 1.

Sakurai, M. 1988. *J.Sol.Chem.*, **17**, 267.

Sakurai, M. 1989. *J.Sol.Chem.*, **18**, 37.

Sakurai, M.; Nakamura, K. and Nitta, K. 1994. *Bull.Chem.Soc.Jpn.*, **67**, 1580.

Sedlbauer, J.; O'Connell, J.P. and Wood, R.H. 2000. *Chem.Geology*, **163**, 43.

Shock E.L. and Helgeson H.C., 1988. *Geochim.Cosmochim.Acta*, **52**, 2009.

Shock, E.L.; Helgeson H.C. and Sverjensky D.A. 1989. *Geochim.Cosmochim.Acta*, **53**, 2157.

Shock E.L. and Helgeson H.C. 1990. *Geochim.Cosmochim.Acta*, **54**, 915.

Shock E.L.; Oelkers E.H.; Johnson J.W.; Sverjensky D.A. and Helgeson H.C. 1992. *J.Chem.Soc.Faraday Trans.*, **88**, 803.

Skold, R.; Suurkuusk, J. and Wadso, I. 1976. *J.Chem.Thermodyn.*, **8**, 1075.

Spink, C.H. and Wadso, I. 1975. *J.Chem.Thermodyn.*, **7**, 561.

Tanger IV, J.C. and Helgeson, H.C. 1988. *Am.J.Sci.*, **288**, 19.

Vliegen, J.; Yperman, J.; Mullens, J.; Francois, J.P. and Van Poucke, L.C., 1984. *J.Sol.Chem.*, **13**, 245.

Wagman, D.D.; Evans, W.H.; Parker, V.B.; Schumm, R.H.; Halow, I.; Bailey, S.M.; Churney, K.L. and Nuttall, R.L. 1982. *J.Phys.Chem.Ref.Data*, **11**: Supplement No. 2.

Wilhelm, E.; Battino, R. and Wilcock, R.J. 1977. *Chem.Review*, **77**, 219.

Wurzburger, S.; Sartorio, R.; Ella,V. and Cascella, C. 1990. *J.Chem.Soc.Faraday Trans.*, **86**, 3891.

Xiao, C.; Bianchi, H. and Tremaine , P.R. 1997. *J.Chem.Thermodyn.*, **29**, 261.

Yasuda, Y.; Tochio, N.M.; Sakurai, M. and Nitta, K. 1998. *J.Chem.Eng. Data*, **43**, 205.

Yezdimer, E. M.; Sedlbauer, J. and Wood, R. H. 2000. *Chem. Geology*, **164**, 259.

PART 3

INVESTIGATION OF CLASSICAL POLARIZABLE WATER MODELS

3.1 Deficiencies of Traditional Non-Polarizable Water Models

As described in part 1, aqueous solutions play a vital role in understanding of chemical evolutionary processes (as well as a large variety of industrial processes). A crucial first step in modeling these systems, from a molecular standpoint, is the development of an accurate potential model of water valid over a wide range of state conditions. Pure water is one of the most difficult systems to accurately model because of its asymmetry, strong polar character and ability to form hydrogen bonds. Since the formation of tertiary ordered structures in large biomolecules is thought to rely heavily on distinctions between polar and non-polar groups, it is imperative that our molecular model for water be able to accuracy describe the density and temperature dependence of water's dielectric and dipole moments.

Traditional intermolecular potential models for water have typically made the assumption of pairwise additivity (Chialvo *et al.*, 1998). While this assumption appears to be rather successful in describing many-body interactions in terms of effective two-body interactions for fluids exhibiting quasi-isotropic polarization effects (Carnie and Patey,1982), it breaks down for systems with anisotropic polarization (Zhu *et al.*, 1994). To account for molecular geometry and dispersion effects, the microscopic description of water has typically been performed through models comprised of a series of fixed-point charges plus non-Coulombic contributions sites (Jorgensen *et al.*, 1983; Watanabe and Klein, 1989; Berendsen *et al.*, 1981; Berendsen *et al.*, 1987). Since these charges cannot change in response to local variations of the molecule's electrostatic environment, they are

unable to describe the changes in the polarization of the medium due to changes in state conditions, presence of surfaces/interfaces, or charged solutes (Israelachvili , 1992). Many-body polarization effects are usually described in simulations by effective pairwise interactions of enhanced (over the dimer) dipole moments. These enhanced dipole moments result from regarding the point charges as adjustable parameters in fitting the model to experimental data. The obvious consequence of this approach is the poor description of either the two-body interactions of the isolated molecular pair (dimer, or system at low density), or/and the many-body interactions of the condensed phase, even when introducing state-dependent force field parameters (Chialvo, 1996).

During the past several years there has been an increasing interest in developing methods for adding various degrees of polarization to already well established simple ridged, fixed charge molecular models of water. These more sophisticated models have the potential of offering a much more accurate description of water, compared to their earlier fixed charged counterparts, because of their explicit inclusion of an anisotropic polarization component. In the following sections we review some of the more common polarizable models and discuss their microstructure and thermodynamics (Chialvo *et al.*, 2000; Yezdimer and Cummings, 1999).

3.2 Review of Polarizable Water Models

Effort to improve the description of water behavior within a classical approach has followed several routes. These strategies include the use of additional interaction sites to handle charge distributions (Jorgensen, 1981), point charge or dipole polarizabilities

(Stillinger and David, 1978; Sprik and Klein, 1988; Rullmann and van Duijnen, 1988; Ahlström *et al*, 1989; Kuwajima and Warshel, 1990; Rick *et al.*, 1994; Dang, 1992; Svishchev *et al.*, 1996). Models employing flexible geometries (Toukan and Rahman, 1985; Lie and Clementi, 1986; Anderson *et al.*, 1987) to account for non-additivity effects have been developed as well as models possessing a combination of flexibility and polarizability (Zhu *et al.*, 1991; Niesar *et al.*, 1990; Famulari, *et al.*, 1998). Unfortunately it would not be practical to discuss all of these models in detail within the confines of this thesis. However for the purposes of a general review, it would be appropriate to describe a few examples of the varying methodologies used to incorporate polarization into a classical molecular model. To these ends we have chosen to review four different polarizable models, namely the SPC-mTR (Liew *et al.*, 1998), the BJH (Bopp *et al.*, 1983), the PPC (Kusalik and Svishchev,1994) and the TIP4P-FQ (Rick *et al.*, 1994) models.

3.2.1 SPC-mTR Model Formulation

The flexible geometry SPC-mTR water model (Liew *et al.*, 1998) is based on equilibrium geometry and non-bonding potentials of the SPC water model (Berendsen *et al.*, 1981). In the SPC monomer geometry each O-H bond is given the unrealistic distance of 1 Å and the HOH angle is taken to be 109.5°. Despite the fact that the SPC geometry is considerably different from the gas-phase experimental value of r_{OH}=0.9752 Å and HOH=104.52°, it is often a favorite starting point for water model developers (although for reasons unknown). The SPC-mTR models also employs a 12-6 Lennard-Jones (LJ)

potential centered on the oxygen site. The LJ parameters and partial charges for the SPC-mTR have been taken to be the same as those from the SPC model. While the partial charges are still held constant, their positions are allow to fluctuate through the introduction of two OH vibrational bonds and a HOH bending potential. The intramolecular binding energy is described by,

$$u_{SPC-mTR}^{intra} = u_{OH} + u_{HOH} \tag{3.1}$$

where the first term on the right hand side accounts for OH stretching and the second term for HOH bending. These terms are defined as,

$$u_{OH} = D_{OH}\left[\left(1-\exp(-\alpha\Delta r_{OH_1})\right)^2 + \left(1-\exp(-\alpha\Delta r_{OH_2})\right)^2\right] \tag{3.2}$$

$$u_{HOH} = \left[L_{rr}\Delta r_{OH_1}\Delta r_{OH_2} + L_{r\theta}(\Delta r_{OH_1} + \Delta r_{OH_2})\Delta r_{H_1H_2} + \frac{1}{2}L_{\theta\theta}\Delta r_{H_1H_2}^2\right]\times \exp[-\beta(\Delta r_{OH_1}^2 + \Delta r_{OH_2}^2)] \tag{3.3}$$

where the adjustable parameters are given by D_{OH}= 426.37 kJ/mol, α=2.511 Å^{-1}, L_{rr}= 651.80 kJ/(mol Å^2), $L_{r\theta}$=-883.01 kJ/(mol A^2), $L_{\theta\theta}$=1107.02 kJ/(mol Å^2), and β=3.0 Å^{-2}. The quantities Δr_{ij} are given by $\Delta r_{ij} = r_{ij} - r_{ij}^{eq}$ where *eq* indicates the SPC (equilibrium) distances. The quantity β is a constraining parameter which localizes the u_{HOH} with the region of r_{OH}<2.5 Å. The final SPC-mTR potential is then described by,

$$u_{SPC-mTR} = 4\varepsilon_{OO}\sum_{i<j}\left\{\left(\frac{\sigma_{OO}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{OO}}{r_{ij}}\right)^{6}\right\} + \sum_{i<j}\sum_{\alpha\neq\beta}\frac{q_{i\alpha}q_{i\beta}}{4\pi\varepsilon_0 r_{i\alpha,j\beta}} + \sum_i (u_{SPC-TR}^{intra})_i \tag{3.4}$$

where the subscripts i,j refer to molecules and the Greek subscripts refer to atoms. Here ε_0 refers to the permittivity of free space and should not be confused with the Lennard-Jones well depth parameter ε_{OO}.

3.2.2 BJH Model Formulation

The BJH water model (Bopp *et al.*, 1983) is modified version of the central model of Stillinger and Rahman (1978). It belongs to a similar class as the SPC-mTR model in that it contains a set of constant point charges connected via an intramolecular vibrational and bending potential. The intramolecular interactions are described using the potential formulated by Carney, Curtiss, and Langhoff and has the following form,

$$u_{BJH}^{intra} = \sum L_{ij}\rho_i\rho_j + \sum L_{ijk}\rho_i\rho_2\rho_3 \tag{3.5}$$

where $\rho_1 = (r_{OH_1} - r_{OH}^{eq})/r_{OH_1}$, $\rho_2 = (r_{OH_2} - r_{OH}^{eq})/r_{OH_2}$, and $\rho_3 = \alpha_{HOH} - \alpha_{HOH}^{eq}$. The quantities r_{OH} and α_{HOH} are the instantaneous bond lengths and angle. The equilibrium values have been taken to be r_{OH}^{eq}=0.9572 Å and α_{HOH}^{eq}=104.52° (the experimental gas phase values). The coefficients, L, have been adjusted to give the best agreement with liquid phase spectrographic results and can be found in Bopp *et al.* (1983).

The intermolecular part of the potential consists of a coulombic term and non-coulombic forces term. The coulombic component is dictated by the partial charges which have been fixed at a constant value of q_H= 0.33 e and q_O = -0.66 e. The non-Coulombic interactions are based on Lemberg and Stillinger's (1975) best empirical functional form for a water monomer and dimer potential's surface and have the following form,

$$u_{OO}^{inter}(r) = \frac{604.6}{r} + \frac{111889}{r^{8.86}} - 1.045\left\{\exp[-4(r-3.4)^2] + \exp[-1.5(r-4.5)^2\right\} \tag{3.6}$$

$$u_{OH}^{inter}(r) = \frac{26.07}{r^{9.2}} - \frac{41.79}{1+\exp[40(r-1.05)]} - \frac{16.74}{1+\exp[5.493(r-2.2)]} \tag{3.7}$$

$$u_{HH}^{inter}(r) = \frac{418.33}{1+\exp[29.9(r-1.968)]} \tag{3.8}$$

where the energies are given in kJ/mol and all distances are given in angstroms.

The final total energy for the BHJ model can then be written as,

$$u_{BHJ} = \sum_{i<j} \sum_{\alpha \neq \beta} \left\{ \frac{q_{i\alpha} q_{j\beta}}{4\pi\varepsilon_0 r_{i\alpha,j\beta}} + u^{inter}_{i\alpha,j\beta}(r_{i\alpha,j\beta}) \right\} + \sum_{i=1} (u^{intra}_{BHJ})_i \tag{3.9}$$

Note that the BHJ model contains van der Waals interactions located on all three atomic sites, instead of only the oxygen (which is typically the case in many classical water models).

3.2.3 PPC Model Formulation

The polarizable point charge (PPC) model of Kusalik *et al.* (Svishchev *et al.*, 1996; Kusalik and Svishchev,1994) retains the simplicity of most non-polarizable three-site models while incorporating the nonadditivity polarization through dynamic point-charges that fluctuate in response to the local electric field. The model consists of a rigid geometry with an OH bond length of 0.943 Å and a HOH angle of 106°. A12-6 Lennard-Jones interaction site is located on the oxygen with σ_{OO}=3.234 Å and ε_{OO} = 0.600 kJ/mol. The partial charges on the hydrogen sites are given by,

$$q_H = 0.486 \pm 0.03E_x + 0.02E_z \tag{3.10}$$

where E_x and E_z are the component of the local electric field (in V/ Å) along the corresponding axis and the sign of the second term changes with the first or second hydrogen. The z-axis for the PPC model is defined to coincide with the monomer's C_2 rotation axis and the x and y-axes are found by defining the entire monomer to lie in the x-

z plane. The magnitude of the oxygen charge is determined via electroneutrality and its position is determined according to the following electric field dependence,

$$\vec{l} = \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} = \begin{pmatrix} -0.025E_x \\ 0 \\ 0.11 - 0.03E_z \end{pmatrix} \tag{3.11}$$

where the vector $\vec{l}$ is the negative charge's location (in Å) with respect to the oxygen Lennard-Jones site. As a result of Eqn. 3.11, the PPC model assumes a planar polarization. The current parameterization predicts a rather high permanent dipole moment (2.14 D) for the isolated water molecule.

In order to assure a self-consistent set of electric field equations an iterative approach is employed. A second order predictor approach, based on three previous self-consistent electric field values, is used to determine the instantaneous electric field around the oxygen site. The second order predictor is,

$$E(t) = 3.0E(t - \Delta t) - 3.0E(t - 2\Delta t) + E(t - 3\Delta t) \tag{3.12}$$

where t is the simulation time and Δt is the length of each time step. Using Eqn. (3.12) one can typically calculate the location of the negative charge within 2-3 iterations to an accuracy of $\delta E^2 / \langle E^2 \rangle < 0.0003$.

3.2.4 TIP4P-FQ Model Formulation

The TIP4P-FQ model was originally developed by Rick and co-workers (1994). The TIP4P-FQ model possesses a geometry that contains three point masses for the oxygen and two hydrogens. The intramolecular distance between the oxygen and

hydrogen is equal to 0.9572 Å and the HOH angle has been fixed at 104.52°. An additional massless point charge lies 0.15 Å along the HOH bisector. The TIP4P-FQ model uses a 12-6 Lennard-Jones potential centered on the oxygen site and intermolecular Coulombic interactions centered on the hydrogen and point charge sites.

The polarization is introduced into the model by allowing the partial charges to fluctuate in response to the local electric field. The method of incorporating fluctuating partial charges is based on the concept of electronegativity equalization (Sanderson, 1951). Parr has shown that the Mulliken electronegativity (χ_i) of an isolated atom is related to the chemical potential of an electron gas (μ_i) surrounding the atom's nucleus,

$$\chi_i = -\frac{\partial E}{\partial N} = -\mu_i = e\frac{\partial E}{\partial Q_i} \tag{3.13}$$

where Q_i is the charge on a atom, E is the ground state energy, N is the number of electrons, and e is an elementary charge. Since the ground state energy is dependent on the presence of an external electrostatic field (introduced by neighboring molecules) the gas-phase partial charge on each atomic site must change to maintain each molecule's electronegativity equilibrium.

The creation/destruction of a partial charge Q_A must be associated with a change in energy since in reality it represents a shift in the electron density of a molecule. The change in energy can be described as a second order expansion of the form,

$$E(Q_A) = E_A(0) + \tilde{\chi}_A Q_A + \frac{1}{2} J^0_{AA} Q_A^2 \tag{3.14}$$

where $E_A(0)$ is the ground state energy of atom A, $\tilde{\chi}_A$ is the Mulliken electronegativity per electronic charge e and J^0_{AA} is twice the hardness of the elecronegativity of the isolated molecule. Thus one can write an expression for the total energy of system of N_{molec} molecules,

$$U = \sum_{i=1}^{N_{molec}} \sum_{\alpha=1}^{N_{atom}} \left[E_\alpha(0) + \tilde{\chi}^0_\alpha Q_{i\alpha} + \frac{1}{2} J^0_{\alpha\alpha} Q^2_{i\alpha} \right] + \sum_{i\alpha<j\beta} J_{\alpha\beta}(r_{i\alpha,j\beta}) Q_{i\alpha} Q_{j\beta} + \sum_{i\alpha<j\beta} V(r_{i\alpha,j,\beta}) \quad (3.15)$$

where r is the interparticle distance. The function $J_{\alpha\beta}(r_{i\alpha,j\beta})$ describes the Coulomb interactions and $V(r_{i\alpha,j,\beta})$ describes any additional non-bonding interactions (*i.e.*, Lennard-Jones). Eqn (3.15) has also been expanded to periodic systems using an Ewald sum (Rick *et al.*, 1994); however we will not go into the details here.

The electronegativity per unit charge of every atom in the system can be given by,

$$\tilde{\chi}_\alpha = \left(\frac{\partial U}{\partial Q_\alpha} \right). \quad (3.16)$$

Eqn (3.16) has the advantage of being easily evaluated at every timestep during a molecular dynamics simulation. In order to ensure conservation of charge during a simulation, a constraint must be introduced into the system. The preferred constraint is that for all *i* molecules,

$$\sum_{\alpha=1}^{N_{atom}} Q_{i\alpha} = 0. \quad (3.17)$$

However other charge constraints are possible (Rick *et al*, 1994). The charge neutralization constraint is introduced into the system by using the method of

undetermined multipliers on the system's Lagrangian. The extended Lagrangian for the TIP4P-FQ molecule is given by

$$L = \sum_{i=1}^{N_{molec}} \sum_{\alpha=1}^{N_{atom}} \frac{1}{2} m_\alpha \left(\frac{dr_{i\alpha}}{dt} \right)^2 + \sum_{i=1}^{N_{molec}} \sum_{\alpha=1}^{N_{atom}} \frac{1}{2} M_Q \left(\frac{dQ_{i\alpha}}{dt} \right)^2 - U - \sum_{i=1}^{N_{molec}} \lambda_i \sum_{\alpha=1}^{N_{atom}} Q_{i\alpha} \tag{3.18}$$

where m_α is mass of an individual atom, M_Q is a fictitious charge mass, λ_i is the Lagrange multipliers, and t is the time. The charge mass as units of energy time^2/charge^2 and is similar in form to the charge mass in a Hoover-Nosé thermostat. The value of M_Q is arbitrary and only affects the efficiency of the charge motion. However one must be careful to avoid coupling between the charge and translation degrees of freedom. Rick *et al.* (1994) have found that M_Q= 6.0*10^{-5} (ps/e)^2 kcal/mol produces essentially no coupling between the two. The atomic degrees of freedom are obtained by integration from Newton's equations of motion,

$$m_\alpha \left(\frac{d^2 r_{i\alpha}}{dt^2} \right) = -\frac{\partial U}{\partial r_{i\alpha}} \tag{3.19}$$

The time evolution of the charge degrees of freedom are obtained by integration of the following relationship,

$$M_Q \left(\frac{d^2 Q_{i\alpha}}{dt^2} \right) = -\frac{\partial U}{\partial Q_{i\alpha}} - \lambda_i = -\tilde{\chi}_{i\alpha} - \lambda_i \tag{3.20}$$

where λ_i can be found to be,

$$\lambda_i = -\frac{1}{N_{atom}} \sum_{\alpha=1}^{N_{atom}} \tilde{\chi}_{i\alpha} \tag{3.21}$$

This allows one to write the equation of motion for the atomic partial charges as,

$$M_Q\left(\frac{d^2Q_{i\alpha}}{dt^2}\right) = -\frac{1}{N_{atom}}\sum_{\beta=1}^{N_{atom}}(\tilde{\chi}_{i\alpha} - \tilde{\chi}_{i\beta}) \tag{3.22}$$

Since the right hand side of Eqn (3.22) can easily be determined at each time step in a simulation, the charges may be propagated using the same numeric integration techniques as typically used with the nuclear motions. In all the calculations presented in this dissertation we have employed a leap-frog Verlet algorithm. The charges are kept near their minimum values (minimizing the electrostatic energy) by using a Hoover-Nosé thermostat (Nosé , 1991) on the fictitious charge temperature (set at ~ 1 K).

The Coulomb interaction denoted by $J_{\alpha\beta}(r_{i\alpha,j\beta})$ was found following the method of Rappé and Goddard (1991). The Coulombic interaction is defined as the Coulomb overlap integral between each atomic site,

$$J_{ij}(r) = \frac{1}{4\pi\varepsilon_0}\int dr_i dr_j \left|\phi_i(r_i)\right|^2 \frac{1}{\left|r_i - r_j - r\right|}\left|\phi_j(r_j)\right|^2 \tag{3.23}$$

where ϕ_i is the appropriate Slater orbital. Values for J^0_{OO} and J^0_{HH} where found to be equal to 1554.8 kJ/(mol e^2) and 1477.0 kJ/(mol e^2) respectively. Since the molecular geometry is fixed the cross intramolecular Coulomb interactions only needed to be evaluated at a fixed distance. The values for $J_{OH}(r_{OH})$ and $J_{HH}(r_{HH})$ were found to be equal to 1198.3 kJ/(mol e^2) and 851.9 kJ/(mol e^2) respectively. At large intermolecular distances, $J_{\alpha\beta}(r_{i\alpha,j\beta})$ approaches the classical limit and allows all intermolecular Coulombic interactions to be approximated by,

$$J_{ij}(r_{ij}) = \frac{1}{4\pi\varepsilon_0 r_{ij}} \tag{3.24}$$

It is preferable to set the isolated gas phase molecule's energy to zero. This can be accomplished through a simple correction to the total energy in Eqn (3.15) for the gas phase energy of each molecule. The correction due to N_{molec} molecules is given by,

$$E_{gas}^{corr} = N_{molec} \frac{(\tilde{\chi}_O^0 - \chi_H^0)^2}{2J_{OO}^0 + J_{HH}^0 - 4J_{OH}(r_{OH}) + J_{HH}(r_{HH})} \tag{3.25}$$

The TIP4P-FQ model was parameterized to the bulk liquid energy, pressure, and pair correlation functions (Soper and Phillips, 1986) at ambient conditions. The paramerization was also constrained to yield the correct gas-phase dipole moment of 1.85 Debye.

3.3 Thermodynamic and Microstructure

For this comparison we have chosen eight representative state conditions, seven of which have been the target of structural analysis using neutron diffraction with isotope substitution (NDIS) by Soper and collaborators (1997) (see Table 3.1). In addition, we also analyze a state condition (number 8 in Table 3.1) which is the only *ab initio* molecular simulation of supercritical water available to date (Fois *et al.*, 1994; Laasonen, *et al.*, 1993; Silvestrelli and Parrinello, 1999). All the simulation results presented in this section were carried out in the NVT ensemble, with N=256 water molecules, using the standard periodic boundary conditions and truncation for the non-electrostatic interactions. Electrostatic interactions were handled through the implementation of Ewald summations, with the exception of the PPC model for which a reaction field method was used.

Table 3.1: State conditions for each water model studied.

State #	Temperature (K)	Density (g/cm^3)
1	298	0.998
2	423	0.998
3	423	0.92
4	573	0.92
5	573	0.77
6	573	0.716
7	673	0.66
8	730	0.64

A quick analysis of the thermodynamic properties indicates a common deficiency, their failure to predict the correct pressure at the conditions where the models were parameterized. Furthermore, a major difference between the two polarizable water models is the large difference in the resulting polarization contribution to the total water dipole moment. In fact, even though both polarizable water models involve two-dimensional polarizabilities (with smaller values of effective values than the corresponding experimental value of $1.44\mathring{A}^3$) the TIP4P-FQ model usually predicts (~50-80%) larger polarization than that of the PPC water model (compare values in Tables 3.2 and 3.3). In contrast, the flexible models exhibit much weaker enhancement of the total dipole (with respect to that of the equilibrium geometry) resulting only from the flexible molecular geometry as opposed to explicit polarizabilities (compare Tables 3.4 and 3.5). Undoubtedly, the observed disparity in the dipolar (polarization) behavior translates into distinct structural features as we will see in the next section. In Figures 3.1-3.3 we compare the predicted behavior of the $g_{OO}(r)$, $g_{OH}(r)$, and $g_{HH}(r)$ by the four water models against the NDIS-97 dataset and the latest *ab initio* molecular dynamics simulation (Silvestrelli and Parrinello, 1999). This comparison clearly indicates that the four models capture the main features of the water structure at ambient conditions (hardly a surprise, since the models are parameterized at these conditions). However, there are still a few noticeable features which are not accounted for by the models, the most obvious being the over-prediction of the strength of the first peak for the three pair correlation functions. It is also interesting to note that even though the *ab initio* and NDIS structure for $g_{OO}(r)$ are in a remarkably good agreement, the *ab initio* simulation still predicts more structured

Table 3.2: Simulation results for the $TIP4P-FQ$ water model ($\mu_{\text{isolated}} = 1.85D$)

State #	P (kbar)	P_{exp} (kbar)	U_c (kJ/mol)[a]	μ (D)	$\mu - \mu_{isolated}$ (D)
1	-0.52	0.01	-41.5	2.62	0.77
2	3.02	1.88	-33.9	2.44	0.59
3	0.95	1.71	-32.4	2.40	0.55
4	4.74	4.89	-27.6	2.31	0.46
5	1.62	0.44	-24.4	2.25	0.40
6	1.02	0.10	-23.1	2.22	0.37
7	1.79	0.80	-19.6	2.16	0.31
8	2.15	1.19	-18.2	2.14	0.29

[a] Configurational energy

Table 3.3: Simulation results for the *PPC* water model ($\mu_{\text{isolated}}(E = 0) = 2.14D$)

State #	P (kbar)	P_{exp} (kbar)	U_c (kJ/mol)[a]	μ (D)	$\mu - \mu_{isolated}$ (D)
1	0.7	0.01	-43.1	2.55	0.41
2	3.2	1.88	-36.6	2.49	0.35
3	1.18	1.71	-35.6	2.48	0.34
4	4.3	4.89	-30.2	2.43	0.29
5	1.6	0.44	-27.6	2.40	0.26
6	1.04	0.10	-26.4	2.38	0.24
7	1.7	0.80	-22.6	2.35	0.21
8	4.3	1.19	-21.0	2.33	0.19

[a] Configurational energy

Table 3.4: Simulation results for the *SPC – mTR* water model ($\mu_{isolated} = 2.27D$)

State #	P (kbar)	P_{exp} (kbar)	U_c (kJ/mol)[a]	μ (D)	$\mu - \mu_{isolated}$ (D)
1	0.23	0.01	-41.5	2.44	0.17
2	0.423	1.88	-33.8	2.42	0.15
3	0.8	1.71	-32.9	2.42	0.15
4	3.37	4.89	-26.1	2.41	0.14
5	0.96	0.44	-23.6	2.41	0.14
6	0.69	0.10	-22.4	2.40	0.13
7	1.2	0.80	-17.5	2.39	0.12
8	1.56	1.19	-15.1	2.39	0.12

[a] Configurational energy

Table 3.5: Simulation results for the *BJH* water model ($\mu_{\text{isolated}} = 1.85D$)

State #	P (kbar)	P_{exp} (kbar)	U_c (kJ/mol)[a]	μ (D)	$\mu - \mu_{isolated}$ (D)
1	13.4	0.01	-40.8	1.973	0.123
2	15.6	1.88	-35.4	1.978	0.128
3	11.6	1.71	-34.5	1.974	0.124
4	14.5	4.89	-29.5	1.979	0.129
5	8.4	0.44	-26.9	1.969	0.119
6	6.7	0.10	-25.6	1.966	0.116
7	6.3	0.80	-21.8	1.964	0.114
8	6.2	1.19	-20.3	1.964	0.114

[a] Configurational energy

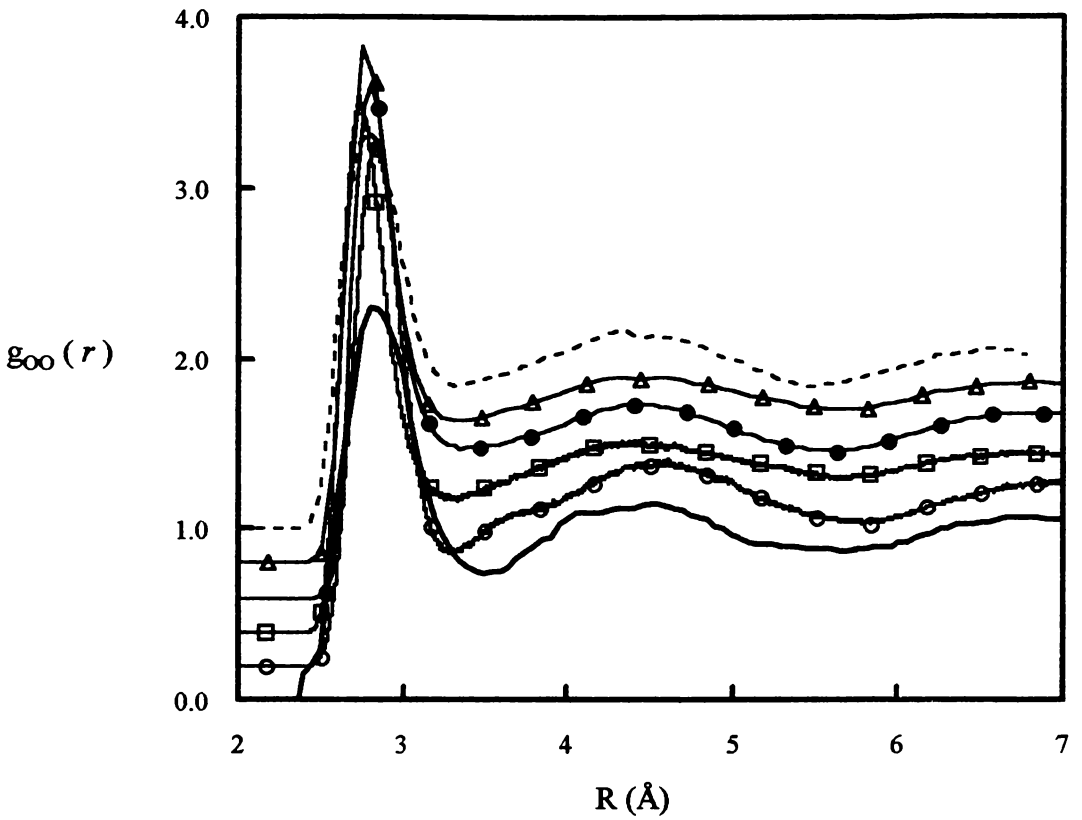

Figure 3.1: Radial distribution function for the O – O interactions. Comparison between the predictions of four water models, an *ab initio* simulation, and the corresponding NDIS results at ambient conditions.

– NDIS-97; --- *ab initio*; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

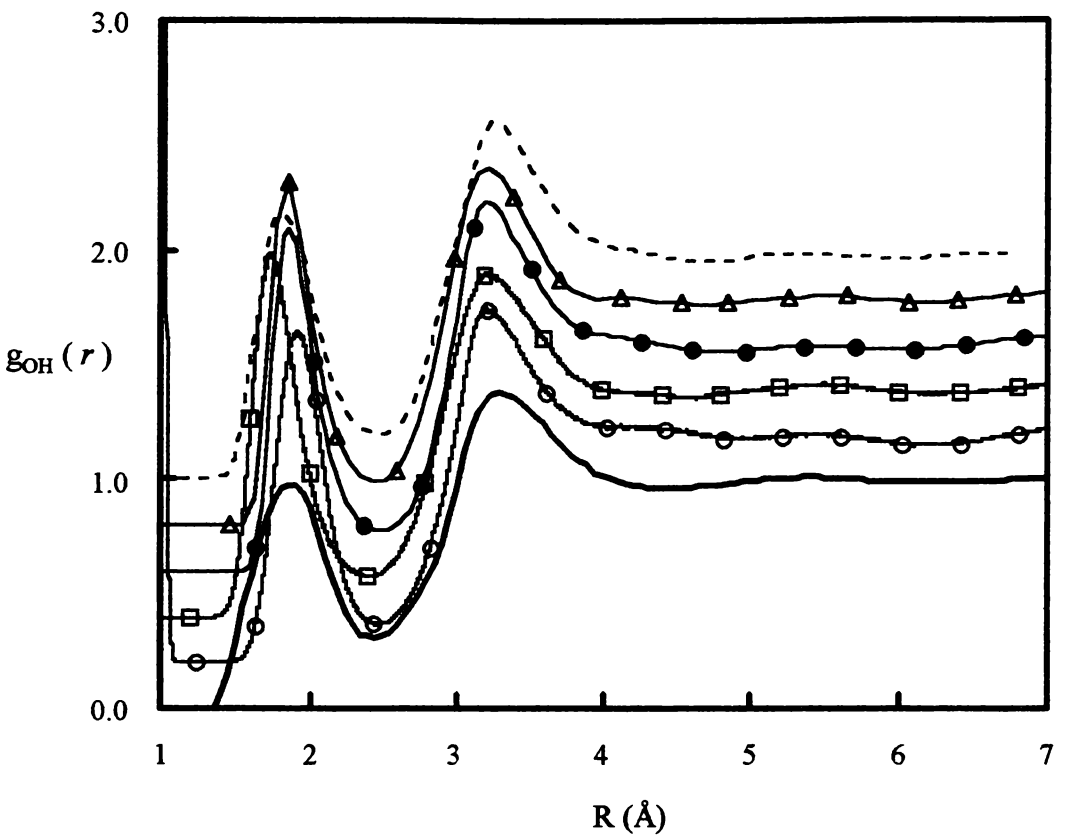

Figure 3.2: Radial distribution function for the O-H interactions. Comparison between the predictions of four water models, an *ab initio* simulation, and the corresponding NDIS results at ambient conditions.

– NDIS-97; --- *ab initio*; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

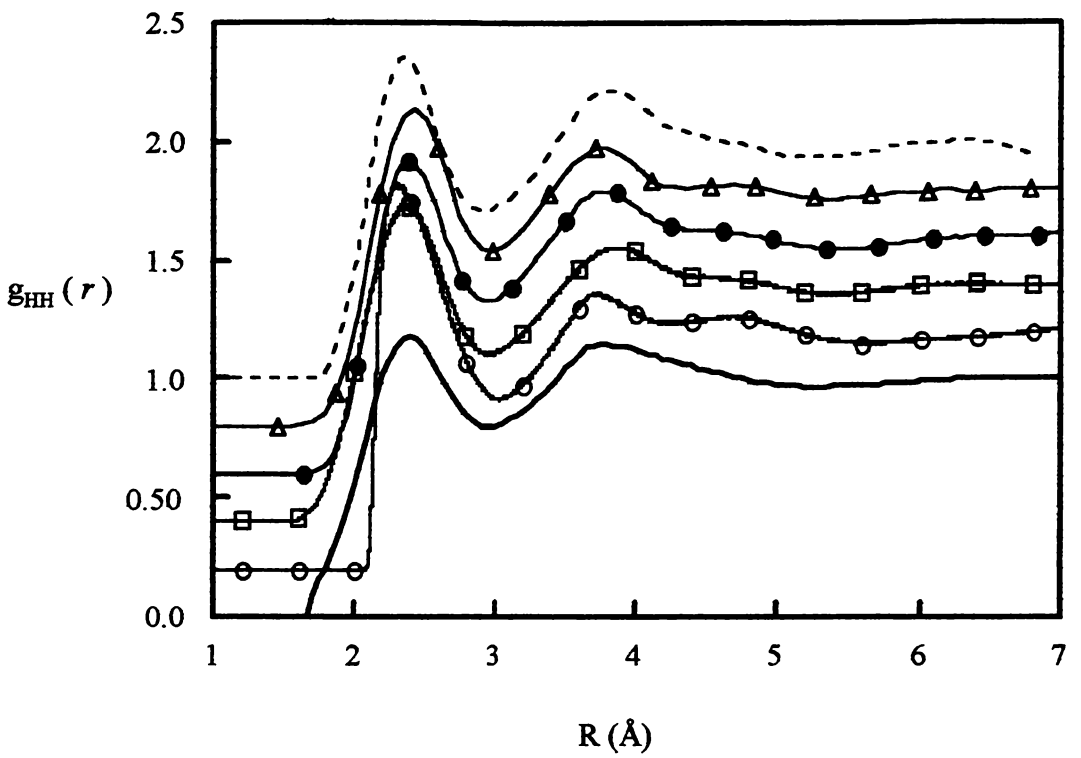

Figure 3.3: Radial distribution function for the H-H interactions. Comparison Between the predictions of four water models, an *ab initio* simulation, and the corresponding NDIS results at ambient conditions.

– NDIS-97; --- *ab initio*; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

peaks for $g_{OH}(r)$, and $g_{HH}(r)$ than the corresponding NDIS (see also Figures 2 and 3 of Silvestrelli and Parrinello, 1999). Moreover, the *ab initio* total dipole moment is $\sim 2.95D$, *i.e.*, at least $0.5D$ larger than those predicted by the polarizable models.

The first set of structures for non-ambient conditions is displayed in Figures 3.4-3.6, *i.e.*, $T = 423K$ and $\rho = 0.998g/cm^3$. The most outstanding feature in this comparison is the disappearing of the second peak of $g_{OO}(r)$, usually associated with the formation of the tetrahedral water structure, and the rather weak $H\cdots H$ correlation in the NDIS results. In contrast, while most models capture the weakening of the $O\cdots O$ correlations (Figure 3.4), the *BJH* model predicts a clear peak at $\sim 4.5\AA$. Moreover, the PPC model displays the closest agreement with NDIS results for $g_{OO}(r)$ and $g_{OH}(r)$. However, neither model can describe the flat structure of the $g_{HH}(r)$ from NDIS, even though (not surprisingly) PPC and TIP4P-FQ give a reasonably good representation.

The second set of structures for non-ambient conditions is shown in Figures 3.7-3.9, *i.e.*, $T = 423K$ and $\rho = 0.92g/cm^3$. The reduction in the system density translates into very small changes in the strength of all site-site correlations. The models predict essentially the same behavior as in the previous case, though they cannot reproduce the shoulder appearing in the experimental $g_{HH}(r)$ at $\sim 2.0\AA$ (see Figure 3.9). In Figures 3.10-3.18 it is possible to analyze how the four models describe the isothermal density effects on the water microstructure. According to Figures 3.10, 3.13, and 3.16 the *PPC* model gives the most accurate description of the behavior of $g_{OO}(r)$, while the *BJH* model predicts the least accurate in that it shows a weak (but noticeable) peak at $\sim 4.5\AA$.

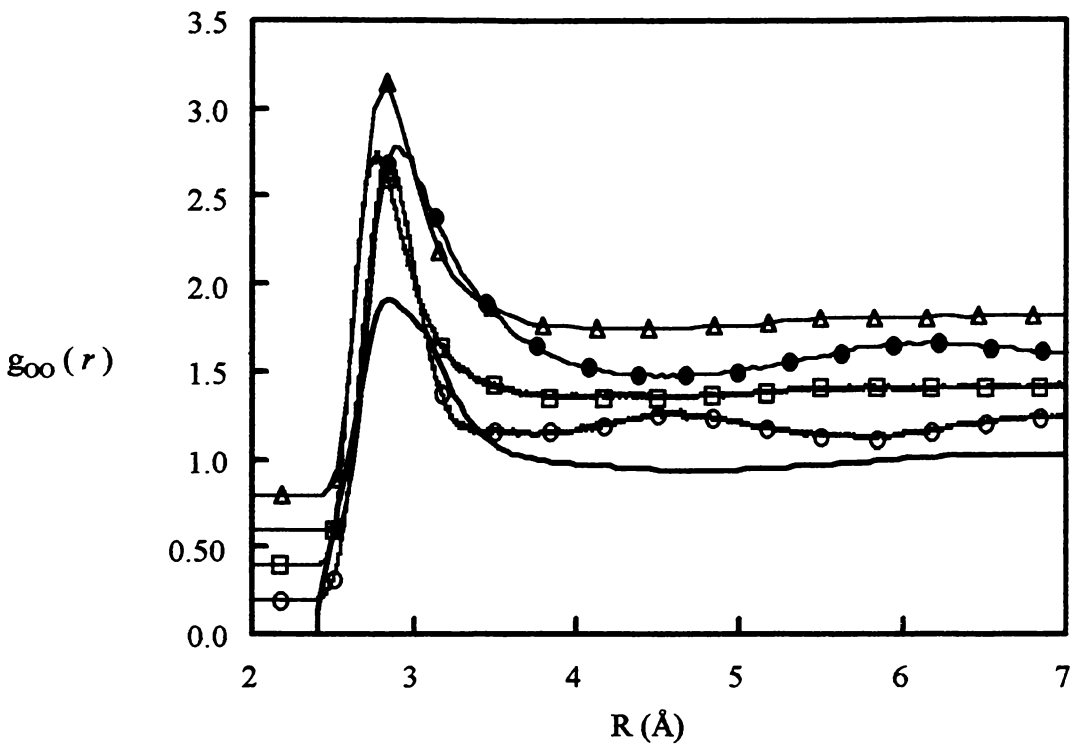

Figure 3.4: Radial distribution function for the O-O interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=423 K and ρ=0.99 g/cm^3.

– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

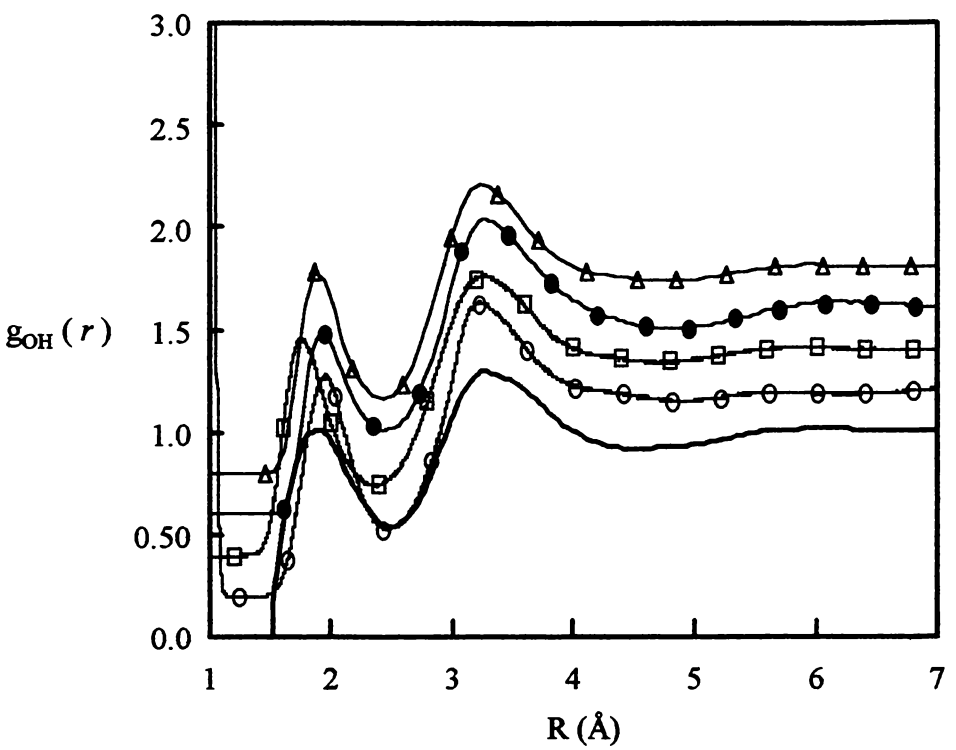

Figure 3.5: Radial distribution function for the O-H interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=423 K and ρ=0.99 g/cm^3.

– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

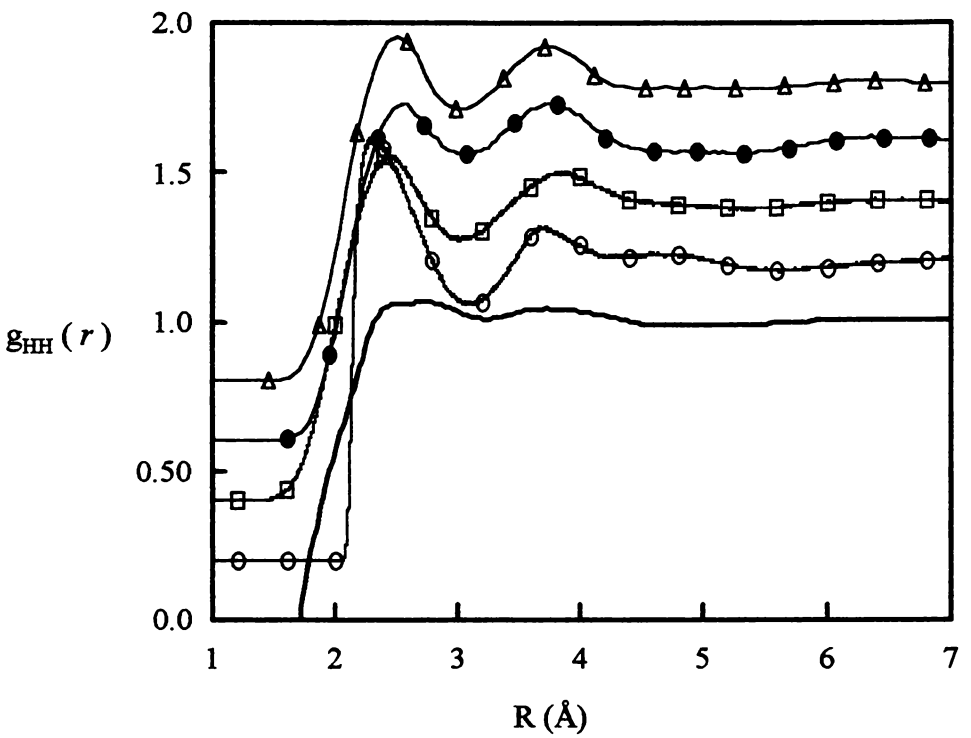

Figure 3.6: Radial distribution function for the H-H interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=423 K and ρ=0.99 g/cm^3.

– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

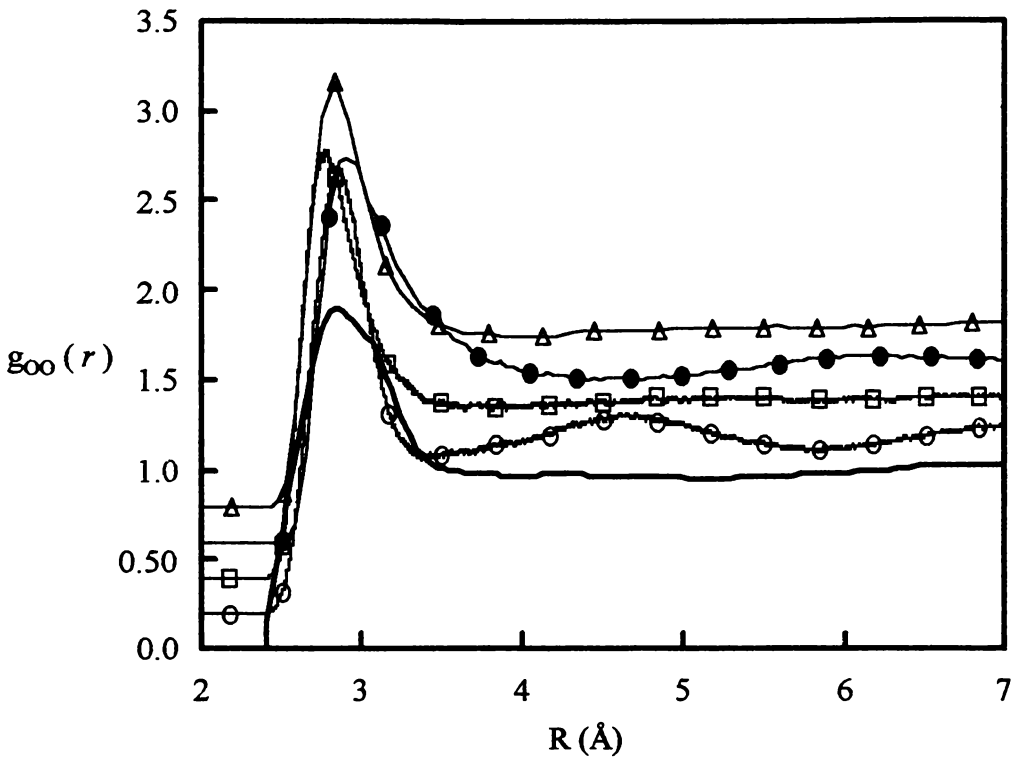

Figure 3.7: Radial distribution function for the O-O interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=423 K and ρ=0.92 g/cm^3.

– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

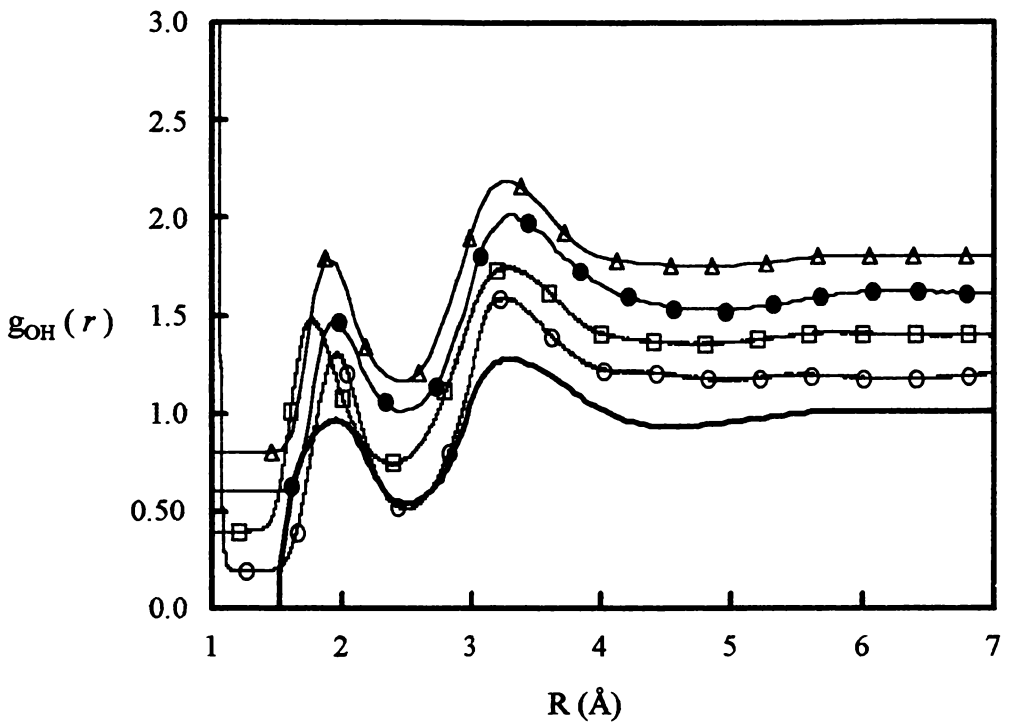

Figure 3.8: Radial distribution function for the O-H interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=423 K and ρ=0.92 g/cm^3.

– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

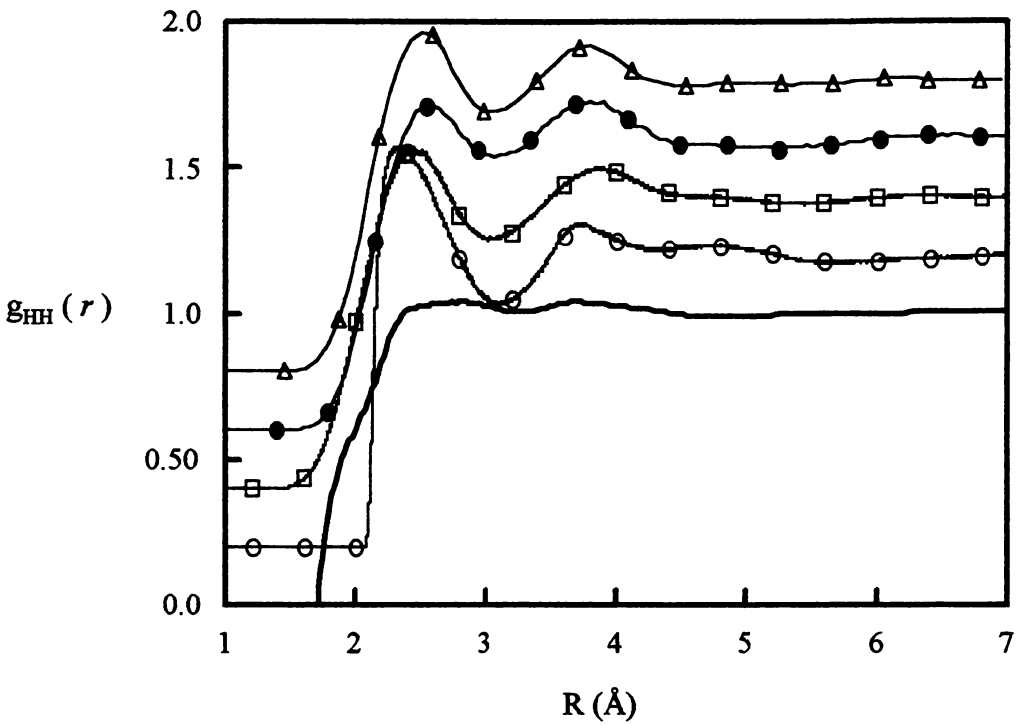

Figure 3.9: Radial distribution function for the H-H interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=423 K and ρ=0.92 g/cm^3.

– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

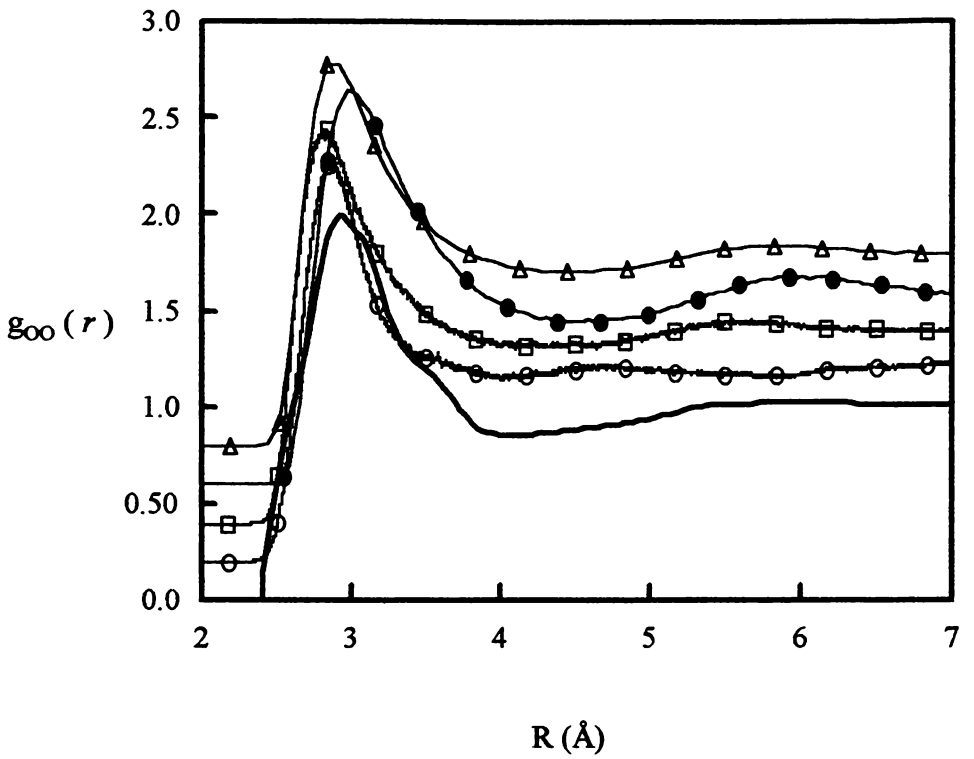

Figure 3.10: Radial distribution function for the O-O interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=573 K and ρ=0.92 g/cm^3.

– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

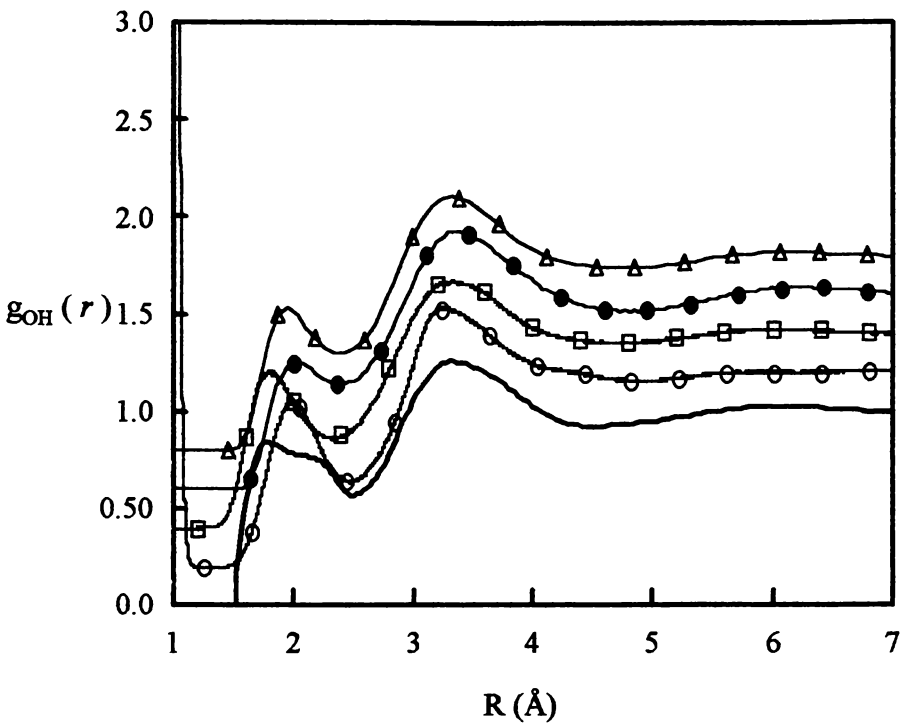

Figure 3.11: Radial distribution function for the O-H interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=573 K and ρ=0.92 g/cm^3.
– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

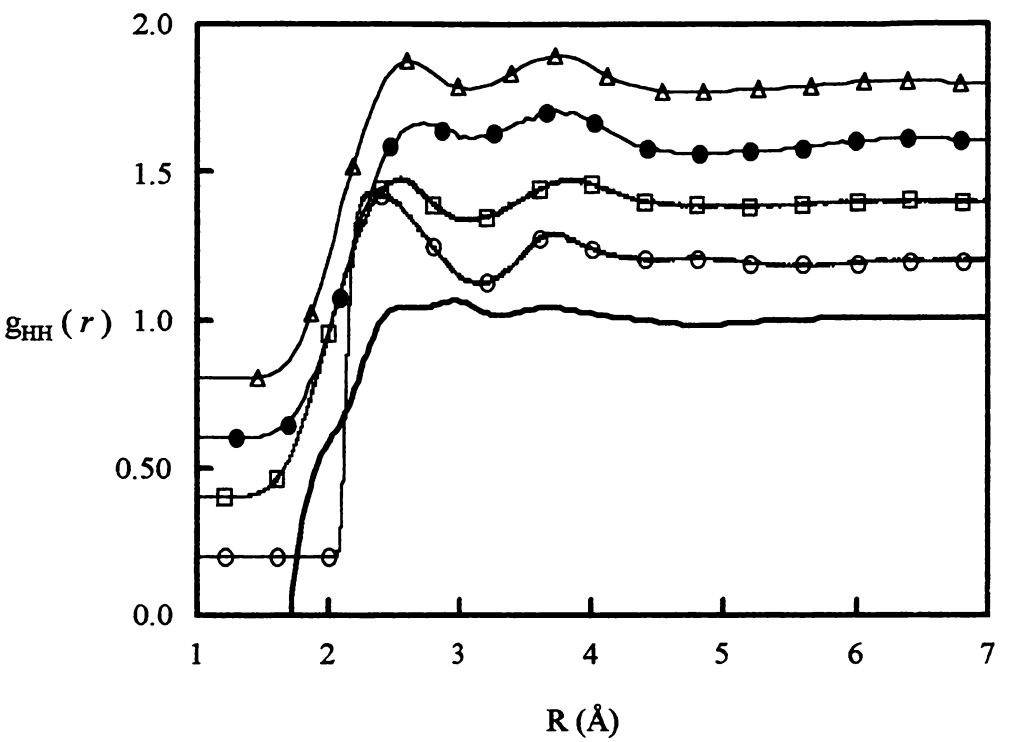

Figure 3.12: Radial distribution function for the H-H interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=573 K and ρ=0.92 g/cm^3.
– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

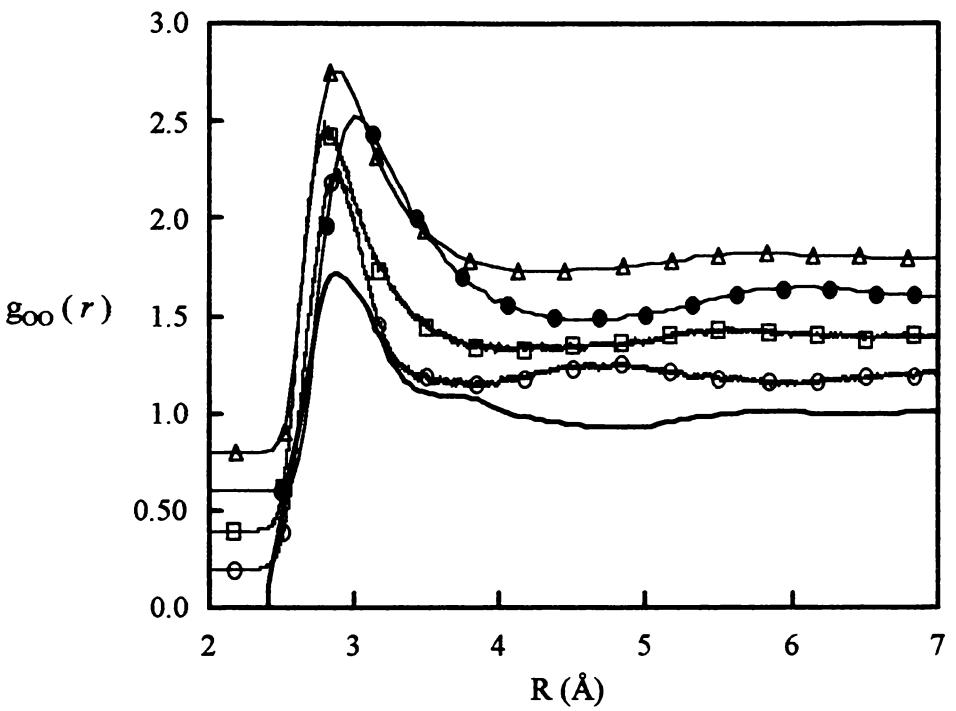

Figure 3.13: Radial distribution function for the O-O interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=573 K and ρ=0.77 g/cm^3.
– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

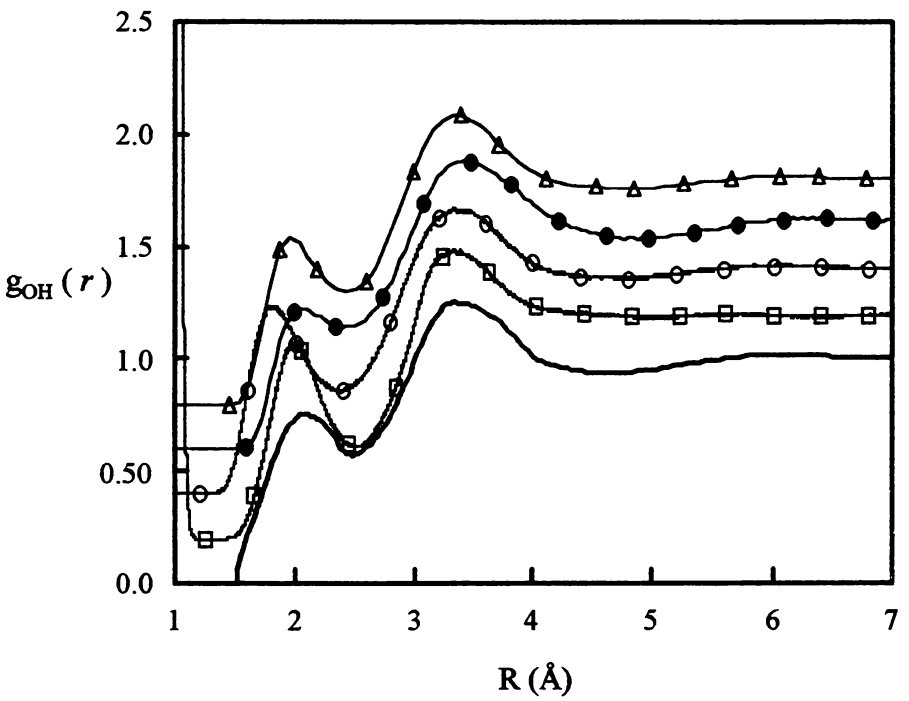

Figure 3.14: Radial distribution function for the O-H interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=573 K and ρ=0.77 g/cm^3.
– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

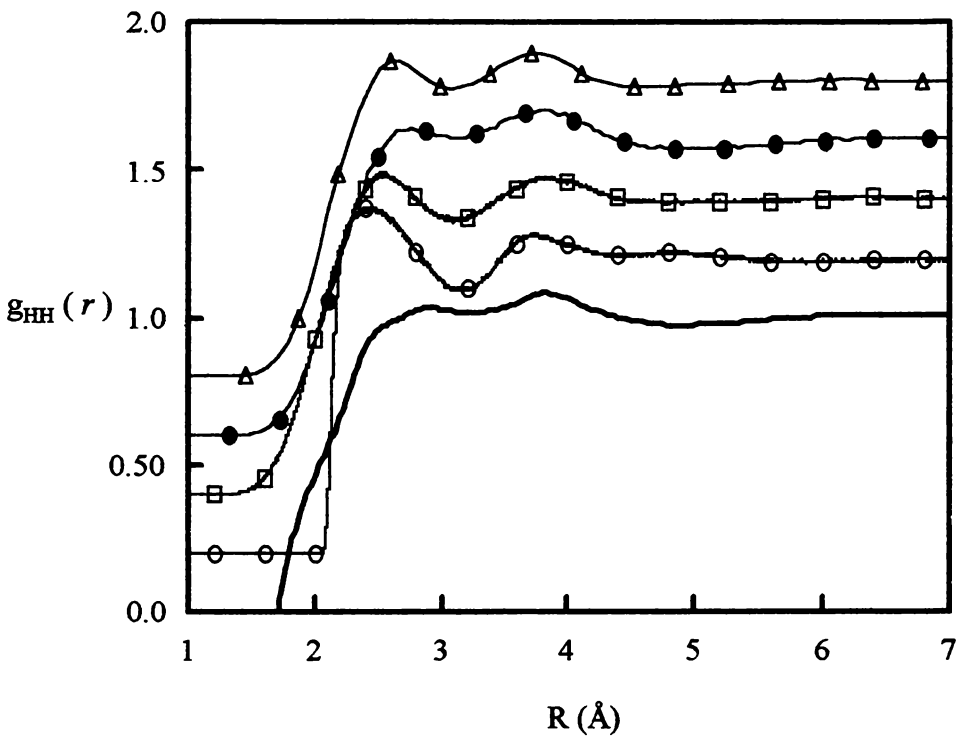

Figure 3.15: Radial distribution function for the H-H interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=573 K and ρ=0.77 g/cm^3.
– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

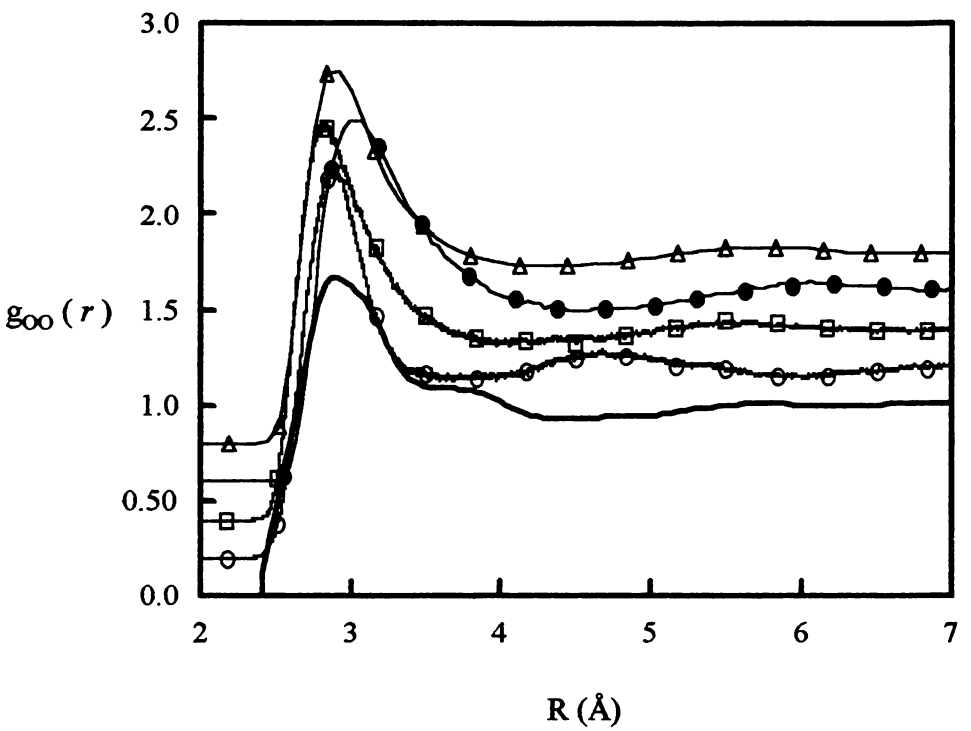

Figure 3.16: Radial distribution function for the O-O interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=573 K and ρ=0.72 g/cm^3.
– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

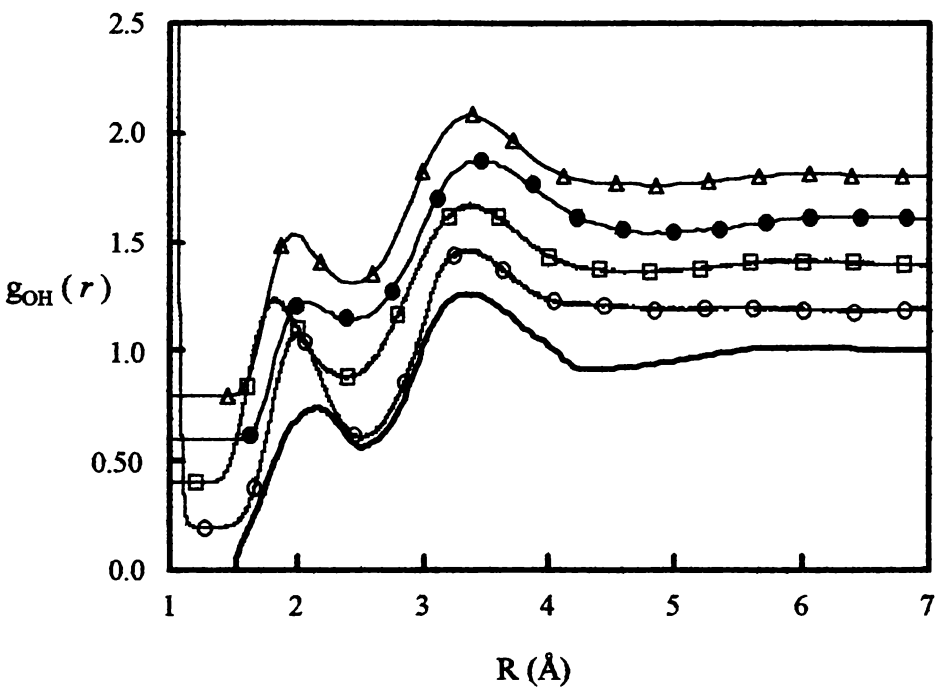

Figure 3.17: Radial distribution function for the O-H interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=573 K and ρ=0.72 g/cm^3.
– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

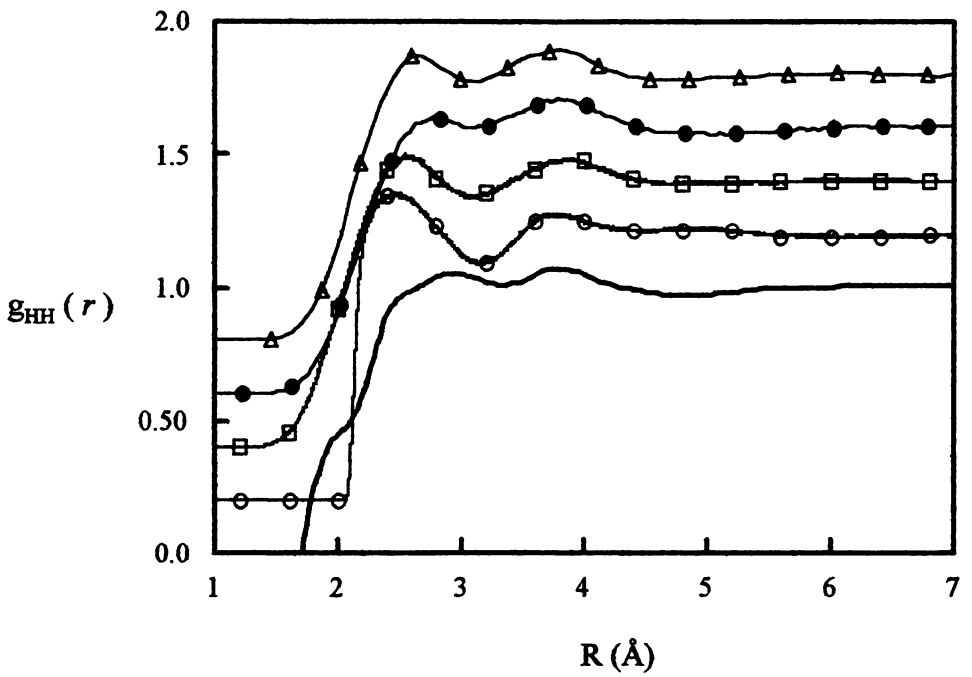

Figure 3.18: Radial distribution function for the H-H interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=573 K and ρ=0.72 g/cm^3.
– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

Note that neither model is able to describe the shoulder on the base of the first peak of $g_{OO}(r)$. Figures 3.11, 3.14, and 3.17 indicate that while all models capture the main features of the $g_{OH}(r)$ (*i.e.*, strength and location of the peaks), none predicts the broadening (shoulder) of the first peak of $g_{OH}(r)$ at $T = 573K$ and $\rho = 0.92g/cm^3$ (Figure 3.11). Moreover, Figures 3.14 and 3.17 clearly indicate the shift to the left (smaller r) of the predicted location of the first peak for the $g_{OH}(r)$ by the SPC-mTR model. In addition, Figures 3.12, 3.15, and 3.18 suggest that the two studied polarizable models give the best representation for the isothermal density dependence of $g_{HH}(r)$, most probably because these models are able to adjust their polarization to the local environment (see $\mu - \mu_{\text{isolated}}(D)$ values in Tables 3.2 and 3.3). Finally, notice that none of the four models reproduces the shoulder in $g_{HH}(r)$ at $\sim 2.0\mathring{A}$.

For the seventh set of state conditions, $T = 673K$ and $\rho = 0.66g/cm^3$, Figures 3.19-3.21 show that TIP4P-FQ evidently gives by far the best agreement with the NDIS results, even when it does not reproduce the shoulders in the first peaks of the corresponding $g_{OO}(r)$, $g_{OH}(r)$, and $g_{HH}(r)$. Note that SPC-mTR predicts again a shift on the location of the first peak of $g_{OO}(r)$ and $g_{OH}(r)$, and that none of the models can describe the deep first valley of $g_{OO}(r)$.

In the final set of state conditions, $T = 730K$ and $\rho = 0.64g/cm^3$, we compare the predictions of the four classical models against an *ab initio* simulation (Figures 3.22-3.24). Even though this first principle simulation involved a very small sample (Fois *et al.*, 1994),

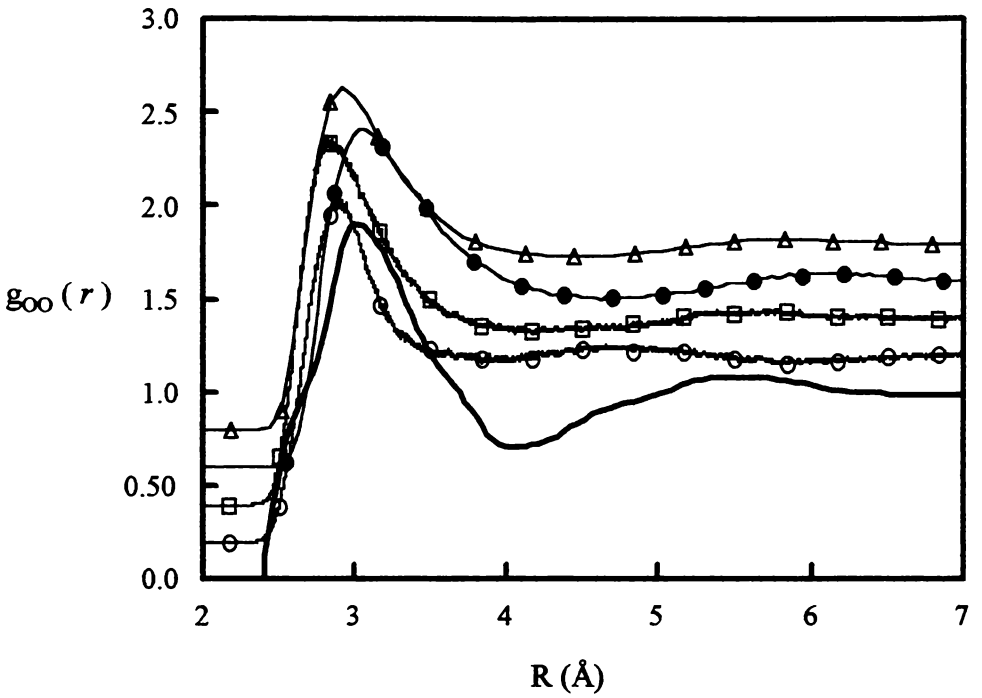

Figure 3.19: Radial distribution function for the O-O interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=673 K and ρ=0.66 g/cm^3.
– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

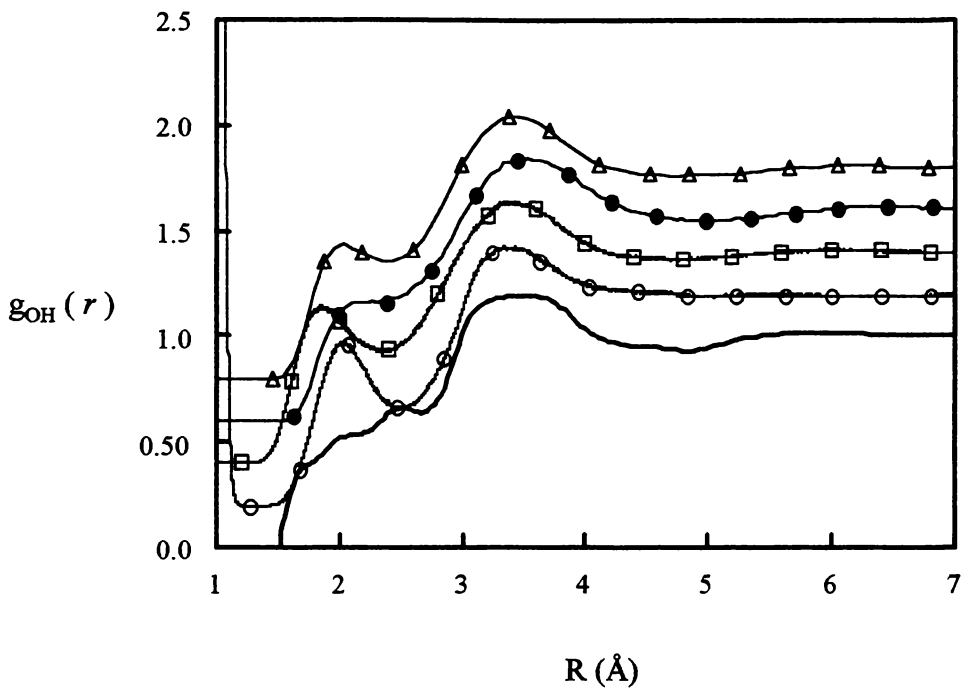

Figure 3.20: Radial distribution function for the O-H interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=673 K and ρ=0.66 g/cm^3.
– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

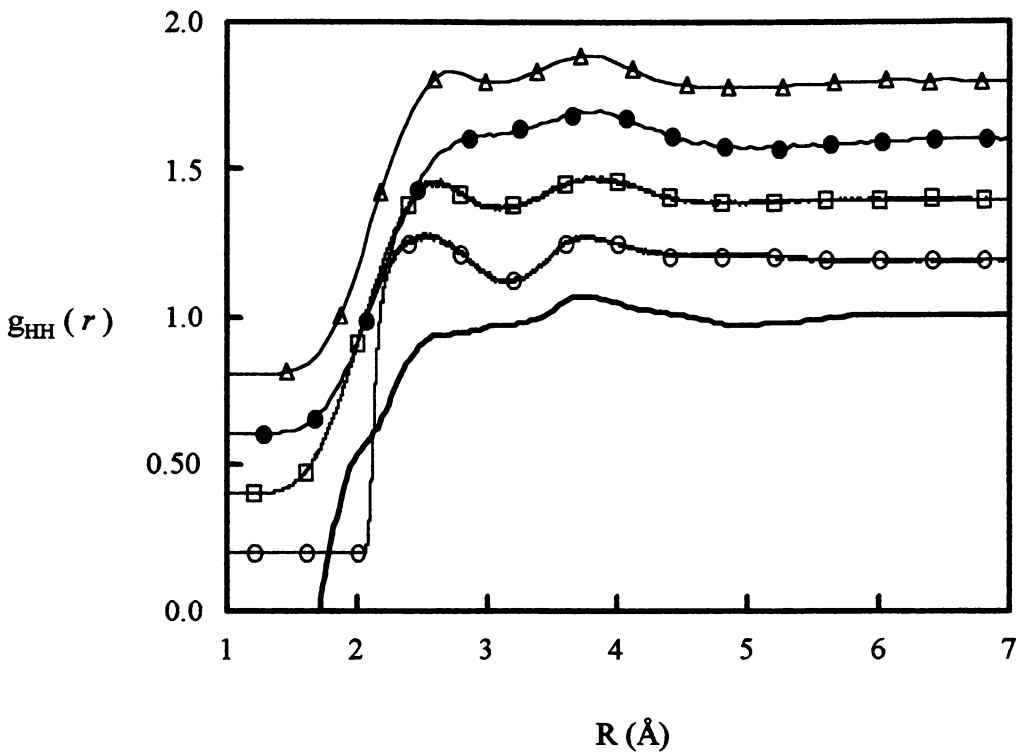

Figure 3.21: Radial distribution function for the H-H interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=673 K and ρ=0.66 g/cm^3.
– NDIS-97; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

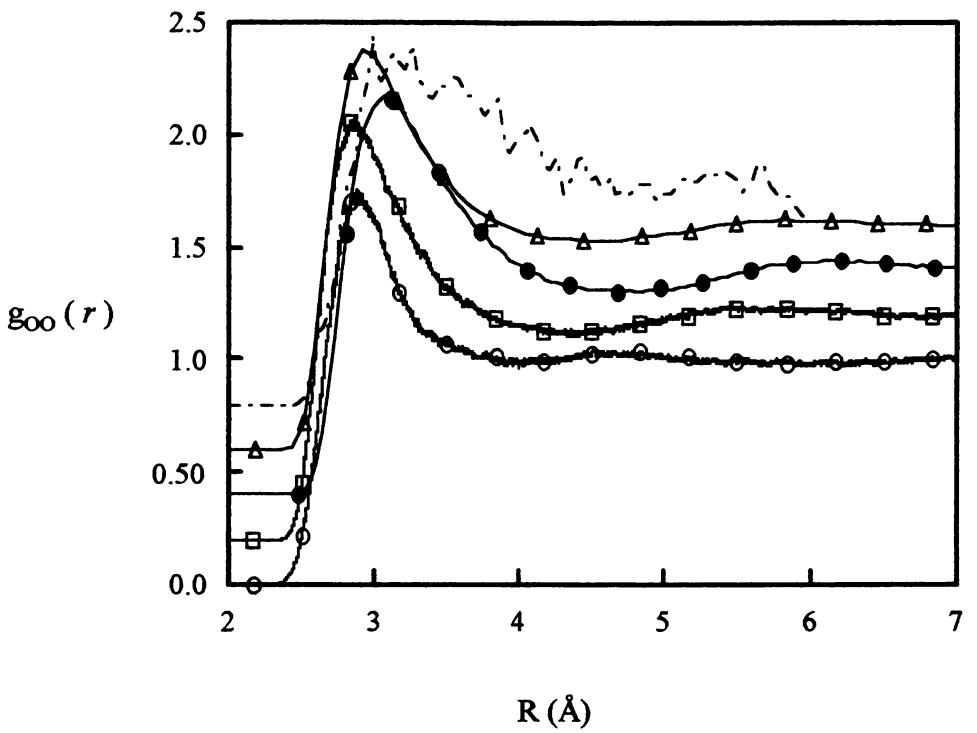

Figure 3.22: Radial distribution function for the O-O interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=730 K and ρ=0.64 g/cm^3.
– NDIS-97; --- *ab initio*; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

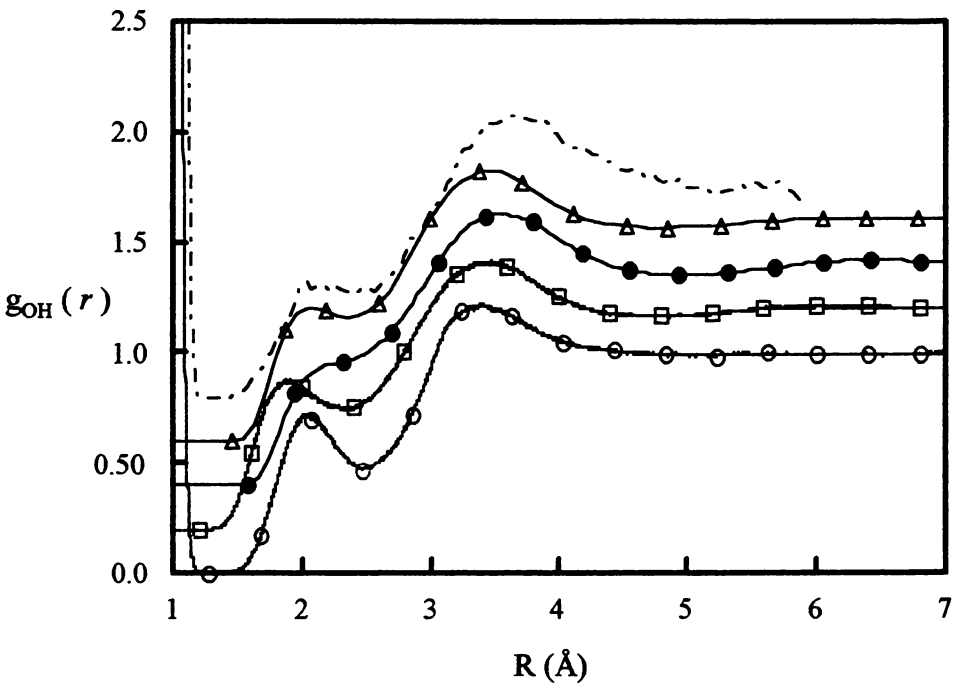

Figure 3.23: Radial distribution function for the O-H interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=730 K and ρ=0.64 g/cm^3.
– NDIS-97; --- *ab initio*; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

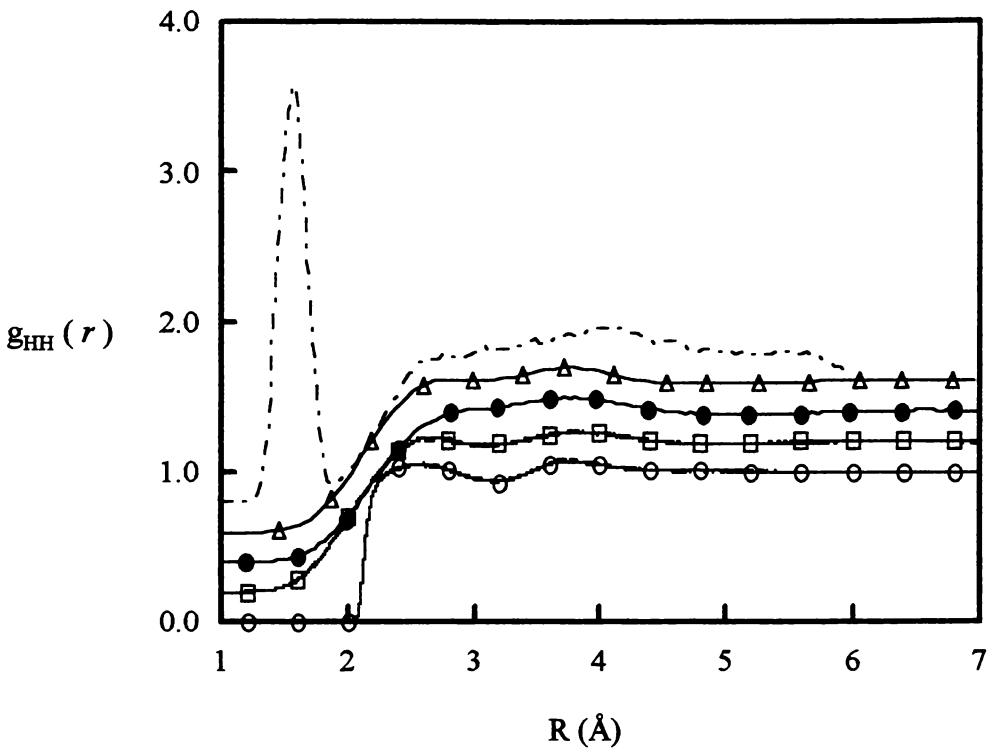

Figure 3.24: Radial distribution function for the H-H interactions. Comparison between the predictions of four water models and the corresponding NDIS results at T=730 K and ρ=0.64 g/cm^3.
– NDIS-97; --- *ab initio*; O-BHJ; □-SPC-mTR; ●- TIP4P-FQ; Δ-PPC

it gives some important structural as well as dipolar information not currently available yet by scattering or other experimental source. Once again, the PPC and TIP4P-FQ models appear to better describe the features of the three correlation functions. For instance these two models describe the first peak of $g_{OH}(r)$ as a shoulder or a flat structure in remarkable agreement with the *ab initio* prediction. Similarly, these polarizable models are able to describe the disappearance of the two first peaks of $g_{HH}(r)$ as a wider or continuous shoulder in the range of $2.0\mathring{A}$ and $4.0\mathring{A}$. These findings are consistent with the fact that the *ab initio* simulation predicts $\mu = (2.3 \pm 0.2)D$ in comparison to 2.33 D and 2.14 D as predicted by the PPC and TIP4P-FQ models. However, the predicted dipole moment from the quantum mechanical simulation much be viewed with caution because the level of theory employed predicts a much higher gas-phase monomer dipole moment than the observed value of 1.85 D.

In conclusion, we find that the fluctuating point charge models (PPC and TIP4P-FQ) perform better than the flexible geometry water models. The PPC model, however, does give a rather poor estimate of the isolated gas phase monomer and its monomer geometry is not consistent with the experimental gas phase geometry. For these reasons we have decided to employ the TIP4P-FQ model for our future investigations at elevated temperatures and pressures, although before we can move forward we must address the TIP4P-FQ model's poor predicted pressure. As mentioned above, all the fluctuating point charge models seem to suffer from this common problem and the exact reason is still unclear. Given this fact, the vapor-liquid phase envelopes for this class of water models could be in considerable disagreement with experiment. To help us better answer these

questions, we have calculated the phase envelope for the TIP4P-FQ water model using a Gibbs-Duhem approach.

3.4 The TIP4P-FQ Liquid-Vapor Phase Envelope

Kiyohara *et al.* (1998) have recently published coexistence data on several different polarizable water models and have concluded that the inclusion of polarization does not improve the description of the phase behavior of water. This result is surprising because one would expect that the explicit addition of a polarizable component to the potential curve would improve the model's description of real water. The phase envelope for the TIP4P-FQ water model was calculated using a Gibbs-Duhem integration technique (Kofke, 1993; 1998). The Gibbs-Duhem equation can be written as

$$d(\beta\mu) = Hd\beta + \beta V dP \quad , \qquad (3.26)$$

where μ, H, V, and P are the chemical potential, molar enthalpy, molar volume, and pressure of the pure substance respectively. The quantity β is equal to $1/kT$, where T is the absolute temperature and k is Boltzmann's constant. Considering a system composed of two coexisting phases, one can easily derive the Clapeyron equation,

$$\left(\frac{dP}{d\beta}\right)_{\substack{saturation \\ curve}} = -\frac{\Delta H}{\beta \Delta V} \qquad (3.27)$$

where ΔH and ΔV are taken to be the difference in molar enthalpy and molar volume of the liquid and vapor phase. Equation (3.27) easily lends itself to numeric integration, provided one has a method of evaluating the right hand side of the equation and at least one initial point on the saturation curve.

Recently, Medeiros and Costas (1997) have determined the vapor-liquid coexistence curve of the TIP4P-FQ model up to 373 K using a Gibbs ensemble Monte Carlo simulation. Taking the initial condition of our numerical integration to be their results at 298 K, Eqn. (3.27) was then integrated using the trapezoidal (first 4 points) and Adams predictor-corrector methods. The inverse temperature step, $\Delta\beta$, of the integration was chosen to be equal to $1.62318^* 10^{-5}$ mol J^{-1}. For each inverse temperature step the predictor estimated the final pressure to within ~ 3.0%. The corresponding corrector was iterated until the pressure had converged to within 0.2 % (or at least 0.02 bar). The corrector steps converged very rapidly, normally requiring only 2-3 iterations at most. The results of our integration are shown in Figure 3.25.

Each calculation of the ΔH and ΔV required in Eqn. (3.27) was found by simultaneously performing two molecular dynamic simulations, one of the liquid phase and one of the vapor phase. MD simulations are a more natural method for calculating the properties of the TIP4P-FQ model than Monte Carlo (MC) methods because its fluctuating molecular charges are treated as an extended Lagrangian. Since the Gibbs-Duhem integration method treats calculations of the liquid and vapor phases as "black boxes", our simulations of TIP4P-FQ can be performed using MD and offers a straightforward method for the calculation of coexisting phases. Since we have initial conditions from the work of Medeiros and Costas (1997), the applications of the Gibbs-Duhem Method is much more efficient than the Gibbs ensemble Monte Carlo method which requires molecule exchanges between the coexisting phases.

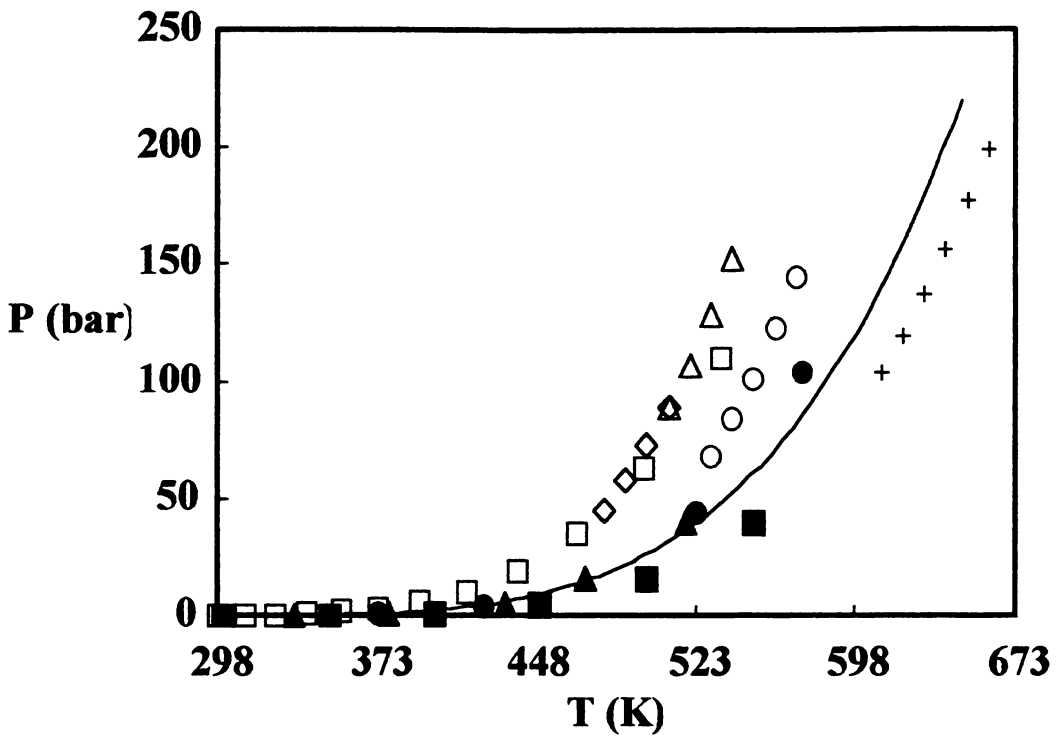

Figure 3.25: Saturation pressure as a function of temperature.

The open symbols represent polarizable models: □ – TIP4P-FQ (this work), O- TIP4P/P, Δ – SPC/P, ◇ –SCPDP, and the + – KJ (Kiyohara, 1998).

The solid symbols represent non-polarizable models: ● – SPC, ■ – SPC/E, and ◆ – MSPC/E (Panagiotopoulos, 1987)

Simulations of the liquid and vapor phases were carried out separately with an NPT ensemble using 256 TIP4P-FQ molecules each. The Newtonian equations of motion were integrated using a 0.5 fs time step and periodic boundary conditions were used. An Ewald sum was also employed to accurately deal with the longer ranged Coulombic interactions. A Hoover-Nosé thermostat (Hoover, 1985; Nosé, 1991) and barostat (Andersen, 1980) were used to maintain constant temperature and pressure throughout the simulations and an additional thermostat using a fictitious mass of $2.5104*10^{-4}$ kJ/mol $(ps/e)^2$ was placed on the total kinetic energy of the fluctuating charges so that the charge temperature remained at about 1 K. This assures that each molecule's point charges remain close to their potential well minima. Each simulation was allowed to equilibrate ~(50 ps – 100 ps) for predictor steps and ~(30 ps – 50 ps) for corrector steps. The values of every H and V were calculated by averaging at least 50 ps. The coexistence curve is shown in Figure 3.26 and is presented in Table 3.6.

The critical temperature and density were estimated by performing a combined least squares fit of the law of recilinear diameters (Frenkel and Smit,1996; Rowlinson, 1959) and the appropriate critical scaling expression. The quantity $(\rho_{liquid} + \rho_{vapor})/2$ was found to contain a slight curvature (see the dashed line in Figure 3.26) and thus the following second order equation for the law of reclinear diameters was used

$$\frac{\rho_{liquid} + \rho_{vapor}}{2} = \rho_c + C_1(T - T_c) + C_2(T - T_c)^2 \quad , \qquad (3.28)$$

where ρ_{liquid} and ρ_{vapor} are the densities of the liquid and vapor phases, ρ_c and T_c are the critical density and temperature, and C_1 and C_2 are adjustable fitting parameters. It is

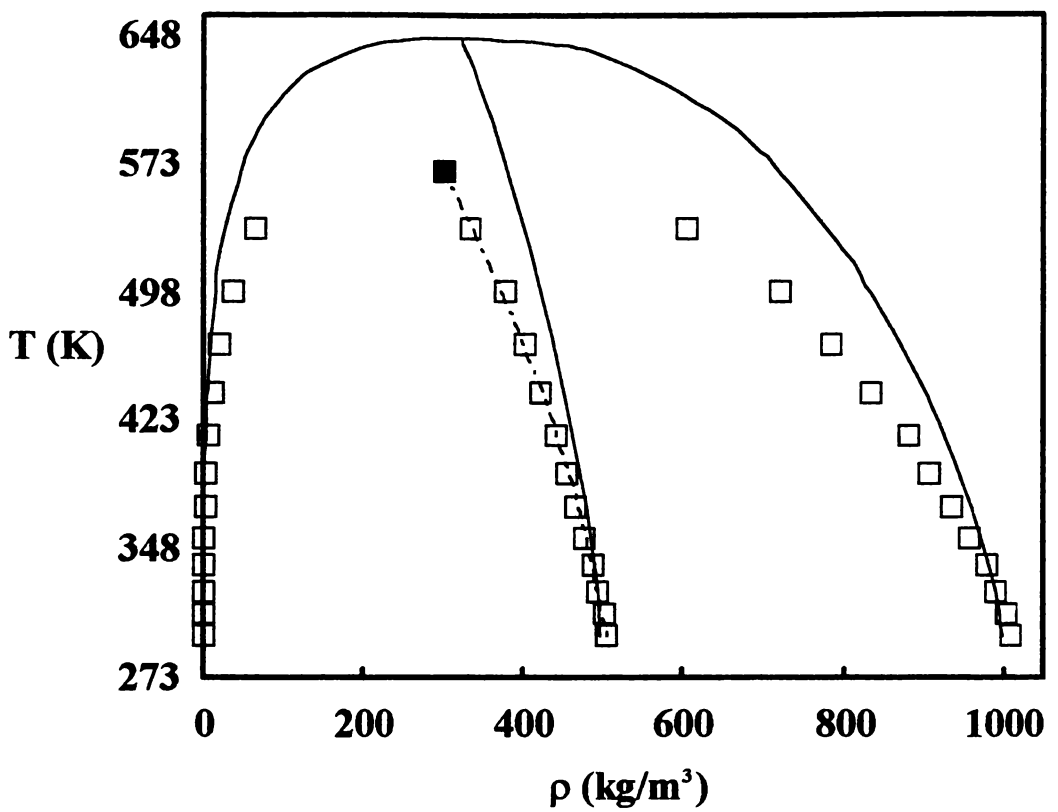

Figure 3.26: Vapor-liquid phase envelope of TIP4P-FQ

The open squares represent our integration points while the solid square represents the value of the critical constants evaluated using Eqn. (3.28) and (3.29). The experimental coexistence curve for water is given by the solid line.

Table 3.6: Vapor-liquid coexistence data for the TIP4P-FQ water model.

T (K)	P (bar)	ρ_{liquid} (kg/m^3)	ρ_{vapor} (kg/m^3)	$\langle D \rangle_{liquid}$ (Deybe)	$\langle D \rangle_{vapor}$ (Deybe)
298.15	0.09397	1010. ± 14	0.07225±0.003	2.63 ± 0.02	1.8619 ± 0.0007
310.65	0.1998	1003. ± 14	0.1339 ±0.04	2.60 ± 0.02	1.8621 ± 0.0009
324.24	0.3926	988.9 ± 17	0.2646 ±0.06	2.57 ± 0.02	1.863 ± 0.001
339.07	0.7647	979.0 ± 15	0.5324 ±0.03	2.53 ± 0.02	1.863 ± 0.001
355.33	1.537	956.8 ± 15	0.9939 ±0.03	2.50 ± 0.02	1.864 ± 0.002
373.24	2.948	933.3 ± 17	1.830 ±0.1	2.46 ± 0.02	1.865 ± 0.002
393.03	5.592	907.7 ± 18	3.222 ±0.2	2.43 ± 0.02	1.868 ± 0.003
415.05	10.32	880.2 ± 20	5.971 ±0.4	2.40 ± 0.02	1.876 ± 0.004
439.67	19.25	833.9 ± 25	11.32 ±0.7	2.35 ± 0.02	1.881 ± 0.005
467.41	35.34	785.6 ± 25	20.03 ±2	2.31 ± 0.02	1.890 ± 0.006
498.88	62.86	720.6 ± 32	36.78 ±3	2.27 ± 0.02	1.914 ± 0.008
534.90	109.8	604.0 ± 44	64.56 ±5	2.20 ± 0.02	1.929 ± 0.01

interesting to note that the experimental data also shows a slight curvature. The equation used to describe the critical scaling was (Kofke, 1993; Rowlinson, 1959),

$$\rho_{liquid} - \rho_{vapor} = C_3(T_c - T)^{0.32} \tag{3.29}$$

where C_3 is another adjustable fitting parameter. All of the data points given in Table 3.6 were used in the fitting of Eqn. (3.28), while only the four highest temperatures were used in Eqn.(3.29). The critical constants were found to be 569.7 ± 1.9 K and 299.7 ± 9 kg/m^3. The values of the remaining adjustable parameters C_1, C_2, and C_3 were found to be 1.14275 kg m^{-3} K^{-1}, -1.418186*10^{-3} kg m^{-3} K^{-2} and 173.898 kg m^{-3} K$^{0.32}$ respectively.

There are three main types of error included in our calculation of the coexistence curve: 1) stochastic errors from our simulations during the calculation of ΔH and ΔV, 2) systematic errors arising from our numerical integration, and 3) errors induced by the initial conditions. The first two can be easily accounted for by various standard statistical methods and are included in our error estimates in Table 3.7. The error introduced by the initial condition is more difficult to determine. While there is no way for us to directly check the accuracy of Medeiros and Costas data without performing a Gibbs ensemble MC calculation (Panagiotopoulos, 1987; 1988), we have confidence in their results because our integration points are in such good agreement (despite the fact we only used their lowest temperature density pair). It is also noteworthy to mention that Meijer and Azhar (1997) have recently developed an interesting method of correcting the coexistence curve for small errors in the initial conditions.

The TIP4P-FQ model under predicts the critical temperature by 13% and the critical density by 7%. This under-prediction of the critical constants is in general

Table 3.7 : Critical constants of several polarizable and non-polarizable water models.

Model	T_c (K)	ρ_c (kg/m^3)
	Polarizable Models	
TIP4P-FQ	569.7 ± 6	299.7 ± 13
TIP4P/P[a]	587 ± 3	350 ± 20
SPC/P[a]	551 ± 1	340 ± 10
SCPDP[a]	538 ± 1	320 ± 10
KJ[a]	685 ± 1	340 ± 10
	Non-Polarizable Models	
SPC[b,c]	593.8 ± 1.2	271 ± 6
SPC/E[b, d]	638.6 ± 1.5	273 ± 9
MSPC/E[b]	609.8 ± 1.1	287 ± 9
Expt.[e]	647.	322

[a] (Kiyohara *et al.*, 1998)
[b] (Panagiotopoulos, 1987)
[c] (DePablo *et al*, 1990)
[d] (Guissani and Guilot, 1993)
[e] (Hill, 1990)

agreement with the findings of Kiyohara *et al.* and suggests that many polarizable models may be suffering from a similar problem. Table 3.7 gives the critical constants of several different polarizable and non-polarizable models. With the exception of the KJ model (Kozack and Jorden, 1992), the non-polarizable models appear to predict the critical constants of water better than polarizable models. Examination of the saturated pressure curve in Figure 3.25 shows that polarizable models are tending to over predict the saturation pressure, while the non-polarizable models (SPC, SPC/E, and MSPC/E) are predicting the pressure curve considerably more accurately. These results are interesting because the dipole moment of both the vapor and liquid phases of the TIP4P-FQ (given in Table 3.6) gives a much better description of the dielectric constant of real water compared to that of the non-polarizable water models. The one exception to this trend appears to be the polarizable KJ model and may be due to the model's exceptionally deep interaction well (Kiyohara *et al.*,1998).

One explanation of this trend could be related to prediction of hydrogen bonding at high temperatures (Guissani and Guilot, 1993). Wood *et al.* (1999) have shown that the TIP4P-FQ has a broader, shallower molecular potential well depth compared to the non-polarizable TIP4P model at ambient conditions. The dimer energy of the TIP4P-FQ model is 3.76 kJ/mol higher than the experimental value (Curtiss *et al.*, 1979) of –22.5 kJ/mol, while the SPC model dimer energy is 5.42 kJ/mol lower than experiment. It appears then that the TIP4P-FQ, and perhaps many polarizable models in general, suffer from a broadening of the potential well introduced from the incorporation of a polarization component into the potential. In addition, Kiyohara *et al.* have shown that small changes

in the potential parameters of the polarizable TIP4P/P model, yield significant improvements in the prediction of water's phase behavior. They have found that small increase in the attractive energy of the TIP4P/P model (1 % reduction in σ_{LJ} ; which leads to an increase in the Coulombic attractive energy) leads to about a 10% increase in the critical temperature. It is therefore reasonable to conclude, that the prediction of liquid-vapor coexistence is much more dependent on the quality of model and parameters describing the first order dependence of the intermolecular interactions, than the method of approximating the second order effect of polarization.

The TIP4P-FQ model does not predict the coexistence curve of water well compared to other simpler non-polarizable models. The model seems to suffer from the same problem as many other polarizable models, namely the under-prediction of the attractive energy due to a broadening of the potential well introduced from the inclusion of polarization. Our findings tend to support the conclusions of Kiyohara *et al.* in that the depth of the potential curve seems to play a larger role in the prediction of the coexistence curve than does the addition of polarization.

As will be seen in part 5, the potential surface of the TIP4P-FQ is in much better agreement with that of high-level *ab-initio* treatment of bulk water than say a non-polarizable model like the TIP4P. Calculations of the dimer energy and dipole moment for a TIP4P-FQ hydrogen bonding pair are in much better agreement with *ab-initio* studies than their corresponding non-polarizable models, such as the TIP4P model. Yet, the reason behind the non-polarizable model's better prediction of the phase co-existence curve is still unclear. Most likely the superiority of the non-polariazble model's co-

existence line over the polarizable model's is somewhat fortuitous, and has only revealed itself due to the appearance of more accurate intermolecular potentials. However, given the complexity involve in modeling such a vexing molecule, the actual cause of the discrepancies in the vapor-liquid equilibrium is most likely not restricted to a single source (*i.e.* the intermolecular potential model). This being said, one area to devote further attention to would be to the treatment of the internal degrees of freedom of the water monomer and, more importantly, to what their possible contribution is on the position of the vapor-liquid coexistence curve (Yezdimer *et al.*, 1999). For instance, the experimental bending frequencies of water in the gas versus the liquid phases can be quite different. Gas phase experiments place the bending modes around 3800-3900 cm^{-1} (Benedict *et al.*, 1956), while liquid phase IR studies have shown them to lie somewhat lower, in the 3300-3500 cm^{-1} range (Maréchal, 1991). This change in the internal motion of a water monomer alone can account for a chemical potential difference of roughly 3-5 kJ/mol between the liquid and vapor phases (vapor minus liquid). Using the equation of state for water developed by Saul and Wagner (1989), it is possible to estimate that a shift of the magnitude in the free energy between the liquid and vapor phases could easily account for increases of up to 70 K in the predicted critical temperature (Figure 3.27).

Indeed, further examination of Figure 3.27 shows that the overall predicted error in the free energy of the liquid phase versus the vapor phase is relatively small, generally less than 5 kJ/mol, supporting the idea that predictions of the liquid-vapor co-existence curve for water can be extremely sensitive to the specifics of the molecular model. Furthermore, given that the typical *ab-initio* errors in the predicted intermolecular energy

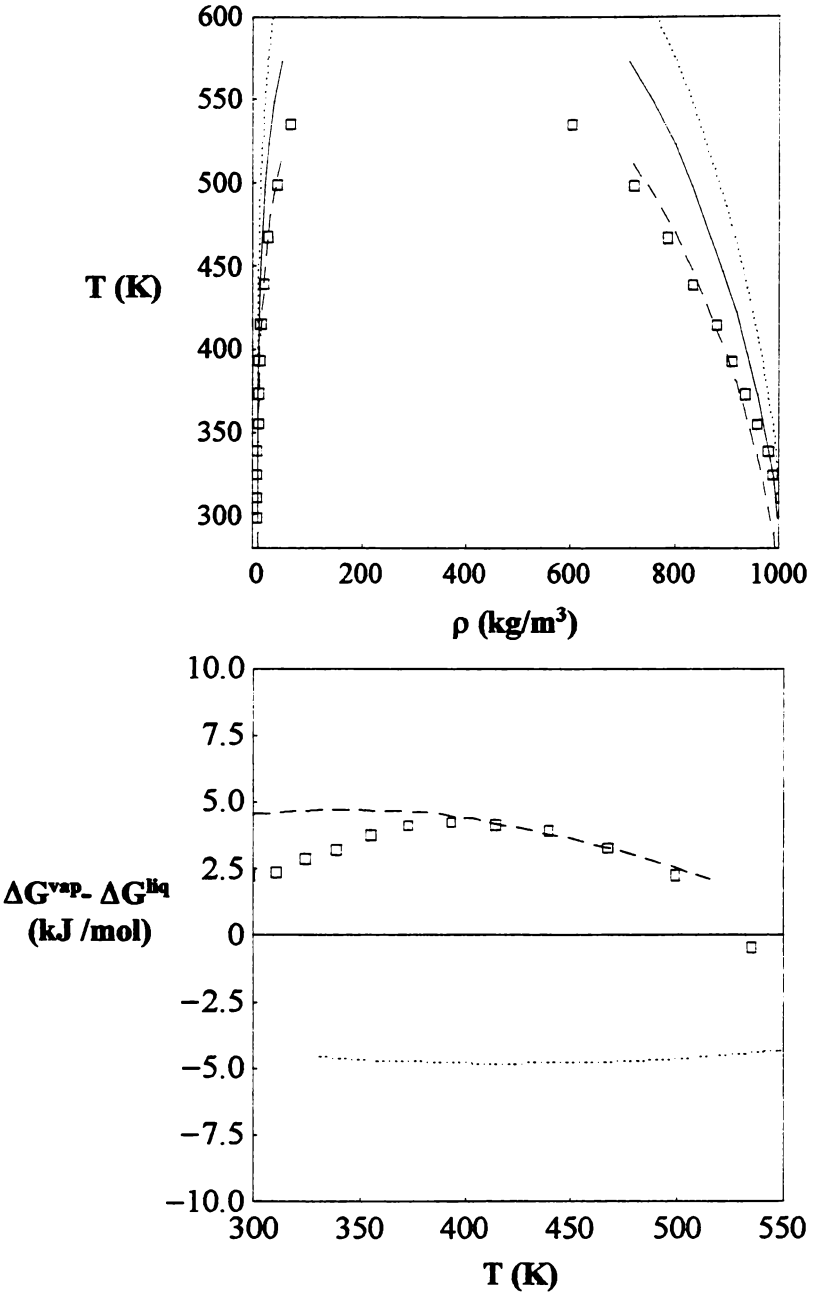

Figure 3.27 : Sensitivity of H_2O phase envelope. The solid line represents the Experimental vapor-liquid coexistence curve. The two dashed lines represent other incorrect phase envelopes and the □ represents the predicted phase envelope of the TIP4P-FQ model. The chemical potential difference between the liquid and vapor phases (along a single temperature tie line) calculated from the Saul and Wanger (1989) EOS is given in the lower graph. One could therefore concluded that an error in experimental value for ΔG^{vap}- ΔG^{liq} of roughly 2.5-5.0 kJ/mol would produce a similar co-existence curve as predicted by the TIP4P-FQ water model.

general occur in the 2-4 kJ/mol range, direct modeling of bulk vapor-liquid equilibrium via *ab-initio* methods seem very difficult at best.

References

Ahlström, P. *et al.* 1989. *Mol. Phys.*, **68**, 563.

Allen, M.P. and Tildesley, D.J. 1987. *The Computer Simulations of Liquids*, Clarendon, Oxford.

Andersen, H.C. 1980. *J. Chem. Phys.*, **72**, 2384.

Anderson, J.; Ullo, J.J. and Yip, S. 1987. *J. Chem. Phys.*, **87**, 1726.

Batista, E.R.; Xantheas, S.S. and Jonsson, H. 1998. *J. Chem. Phys.*, **109**, 4546.

Benedict, W.S.; Gailar, N. and Plyer, E.K. 1956. *J. Chem. Phys.*, **24**, 1139.

Berendsen, H.J.C.; Postma, J.P.M.; von Gunsteren, W.F. and Hermans, J. 1981. *Intermolecular Forces*, edited by Pullman,B. (Reidel, Dordrecht), 331.

Berendsen, H.J.C. *et al.* 1984. *J. Chem. Phys.*, **81**, 3684.

Berendsen, H.J.C.; Grigera, J.R.and Straatsma, T.P. 1987. *J. Phys. Chem.*, **91**, 6269.

Beveridge, D.L. *et al.* 1983. *Monte Carlo simulation studies of the equilibrium properties and structure of liquid water*, in *Molecular-Based Study of Fluids*, J.M. Haile and G.A. Mansoori, Editors., American Chemical Society: Washington, D.C.

Bopp, P.; Jancsó, G. and Heinziger,K. 1983. *Chem. Phys. Lett.*, **98**, 129.

Boulougouris, G.C.; Economou, I.G. and Theodorou, D.N. 1998. *J. Phys. Chem. B.*, **102**, 1029.

Carnie, S.L. and Patey, G.N. 1982. *Mol. Phys.*, **47**, 1129.

Chialvo, A.A. and Cummings, P.T. 1996. *J. Chem. Phys.*, **105** , 8274.

Chialvo, A.A.; Yezdimer, E.M.; Driesner, T.; Cummings, P.T. and Simonson, J.M. 2000. *Chem. Phys.*, **258**, 109.

Chialvo, A.A. *et al.* 1998. *Ind. Eng. Chem. Res.*, **37**, 3021.

Chialvo, A.A., 1996. *J. Chem. Phys.*, **104**, 5240.

Coulson, C.A. and Eisenberg, D. 1966. *Proc. Royal Soc. London*, **A291**, 445.

Curtiss, L.A.; Frurip, D.J. and Blander, M. 1979. *J. Chem. Phys.*, **71**, 2703.

Dang, L.X. 1992. *J. Chem. Phys.*, **97**, 2659.

De Pablo, J.J.; Prausnitz , J.M.; Strauch, I.I. J. and Cummings, P.T. 1990. *J. Chem. Phys.*, **93**, 7355.

Famulari, A. *et al.*, 1998. *J. Chem. Phys.*, **108**, 3296.

Frenkel, I.D. and Smit, B. 1996. *Understanding Molecular Simulation* (New York : Academic Press).

Fois, E.S.; Sprik, M. and Parrinello, M. 1994. *Chem. Phys. Lett.*, **223**, 411.

Gatti, C.; Silvi, B. and Colonna, F. 1999. *Chem. Phys. Lett.*, **247**, 135.

Gregory, J.K. *et al.* 1997. *Science*, **275**, 814.

Guissani, Y. and Guilot, B. 1993. *J. Chem. Phys.*, **98**, 8221.

Hill, P.G., 1990. *J. Phys.Chem.Ref.Data* , **19**, 1233.

Hoover, W.G., 1985. *Phys. Rev.*, **A31**, 1695.

Israelachvili, J. 1992. *Intermolecular & Surface Forces*, London: Academic Press.

Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W. and Klein, M.L. 1983, *J. Chem. Phys.*, **79** , 926.

Jorgensen, W.L., 1981. *J. Am. Chem. Soc.*, **103**, 335.

Kozack, R.E. and Jordan, P.C., 1992, *J. Chem. Phys.*, **96**, 3120.

Kusalik, P.G. and Svishchev, I.M. 1994. *Science*, **265**, 1219.

Kuwajima, S. and Warshel, A. 1990. *J. Phys. Chem.*, **94**, 460.

Kiyohara, K.; Gubbins, K.E. and Panagiotopoulos, A.Z. 1998. *Mol. Phys.*, **94**, 803.

Kofke, D.A. 1993. *J.Chem. Phys.*, **98**, 4149.

Kofke, D.A. 1998, *Adv. Chem. Phys.*, **105**, 405.

Laasonen, K.; Sprik, M. and Parrinello, M. 1993. *J. Chem. Phys.*, **99**, 9080.

Lemberg, H.L. and Stillinger, F.H. 1975. *J. Chem. Phys.*, **62**, 1677.

Lie, G.C. and Clementi, E. 1986. *Phys. Rev. A*, **33**, 2679.

Liew, C.C.; Inomata, H. and Arai, K. 1998. *Fluid Phase Equil.*, **144**, 28798.

Maréchal, Y. 1991. *J. Chem. Phys.*, **95**, 5565.

Medeiros, M. and Costas, M.E. 1997. *J. Chem. Phys.*, **107**, 2012.

Meijer, E.J. and Azhar, F.E. 1997. *J. Chem. Phys.*, **106**, 4678.

Niesar, U. *et al.* 1990. *J. Phys. Chem.*, **94**, 7949.

Nosé, S. 1991. *Progress of Theoretical Physics Supplement*, **103**, 1.

Parr, R.G. and Yang, W. 1989. *Density-Functional Theory of Atoms and Molecules* Oxford University , Oxford.

Rappé A. K. and Goddard W.A. 1991. *J. Phys. Chem.*, **95**, 3358.

Rick, S.W.; Stuart, S.J. and Berne, B.J. 1994. *J. Chem. Phys.*, **101**, 6141.

Rullmann, J.A.C. and van Duijnen, P.T. 1988. *Mol. Phys.*, **63**, 451.

Rowlinson, J.S. 1959. *Liquids and Liquid Mixtures*, Butterworths Scientic publications, London.

Panagiotopoulos, A.Z. 1988. *Mol. Phys.*, **63** , 527.

Panagiotopoulos, A.Z. 1987. *Mol. Phys.*, **61** , 813.

Sanderson, R. T. 1951. *Science*, **114**, 670.

Saul, A. and Wagner, W. 1989. *J. Phys. Chem. Ref. Data*, **18**, 1537.

Silvestrelli, P.L. and Parrinello, M. 1999. *J. Chem. Phys.*, **111**, 3572.

Stillinger, F.H. and David, C.W. 1978. *J. Chem. Phys.*, **69**, 1473.

Stillinger, F.H. and Rahman, A. 1978. *J. Chem. Phys.*, **68**, 666.

Soper, A.K. and Phillips, M.G. 1986. *Chem. Phys.* **107**, 47.

Soper, A.K.; Bruni, F. and Ricci, M.A. 1997. *J. Chem. Phys.*, **106**, 247.

Sprik, M. and Klein, M.L. 1988. *J. Chem. Phys.*, **89**, 7556.

Svishchev, I.M.; Kusalik, P.G.; Wang, J. and Boyd, R.J. 1996. *J. Chem. Phys.*, **105**, 4742.

Watanabe, K. and Klein, M.L. 1989. *Chem. Phys.*, **131**, 157.

Wood, R.H.; Yezdimer, E.M.; Sakane,S.; Barriocanal, J.A. and Doren, D.J. 1998, *J. Chem. Phys.*, **110** , 1329.

Yezdimer, E.M.; Chialvo, A.A. and Cummings, P.T. 1999. AIChe Annual Meeting, Dallas, TX.

Zhu, S. B.; Singh, S. and Robinson, G. W. 1994. *Field-perturbed Water*, in *Modern Nonlinear Optics*, M. Evans and S. Kielich, Editors., John Wiley & Sons, Inc.: New York.

Zhu, S. B.; Singh, S. and Robinson, G. W. 1991. *J. Chem. Phys.*, **95**, 2791.

PART 4

EXAMINATION OF THE FREE ENERGY PREDICITIONS IN THE NEAR CRITICAL REGION

4.1 Possible Implications for Alterative Industrial Solvents

Besides their possible role in the initial chemical evolution of life, hydrothermal solutions also may possess several industrially beneficial properties. Hydrothermal solutions are generally defined as dilute aqueous mixtures at $T > 373$ K and because of the very polar nature of water their solvating properties can vary greatly as a function of temperature and pressure. Since there has been considerable evidence that nature may have utilized these solutions in the creation of the organic molecules necessary for the emergence of life, perhaps modern industry could exploit them to increase the efficiency of its own manufacturing efforts.

In particular, supercritical fluids offer considerable promise for serving as alternative, more environmentally friendly industrial solvents. In general, supercritical fluids (SCF) possess physical properties (*e.g. diffusion and viscosity*) that typically lie between the liquid and gas phase values. Because of the high compressibilities characteristic of SCFs, large variations in these properties can be introduced via modest changes in temperature and/or pressure, thus serving to create a unique solvent environment (Chialvo and Cummings, 1999). In particular, the ability of SCFs to dissolve a wide variety of both polar and non-polar species makes them ideal candidates for new industrial solvents and may allow for previously infeasible synthetic processes (*e.g. organometallic*) to be utilized (Savage *et al.*, 1995).

Unfortunately, the use of the ideal alternative solvent, water, is plagued by two major problems at supercritical conditions: ion association and supercritical water oxidation (SCWO). At supercritical conditions, the reduced dielectric constant of water

promotes ion pairing and reactions with reactor walls resulting in a high degree of corrosion (Mitton *et al.*, 1995). In addition unwanted side reactions can occur with oxygen impurities and even water itself, yielding a significant amount of decomposition of the organic reactants and/or products (Brock *et al*, 1996; Yu and Savage, 1998).

Over the past several years there have been considerable advancements made in the development of equations of state (EOS) for infinitely dilute aqueous organics at temperatures above 373 K (Amend and Helgeson, 1997; Shock and Helgeson, 1990; Plyasunov *et al.*, 2000). Some most recent correlation work, described in part 2, was carried out by Yezdimer *et al.* (2000) and makes use of an equation of state based on Kirkwood and Buff's theory of solution (Kirkwood and Buff, 1951; Sedlbauer *et al.*, 2000). The approach followed by Yezdimer *et al.* centers on the integration of the partial molar volumes to obtain the residual chemical potentials (O'Connell, 1994) and makes use of a simple functional group addivity scheme to ensure the widest range of solute compositions. This study, however, is limited to acyclic organic compounds due to the lack of experimental data at temperatures above 400 K.

While the equation of Yezdimer *et al.* performs quite well at temperatures under 550 K, the predictions for several different classes of acyclic organic compounds show an unexpected shift in their predicted free energies of hydration in the near-critical region of the solvent ($T > 620$ K; P=28 MPa). These predicted shifts in the free energies of hydration would imply that larger, longer chain molecules (such as hexane) should have higher solubilities than smaller, shorter chain molecules (such as propane or butane). This behavior is also not restricted to alkanes. Shifts in the solubility for series of acyclic alcohols, amides, and amines are also predicted to occur (Yezdimer *et al.*, 2000).

At the temperatures and pressures where this predicted effect begins pure water typically has a density on the order of ~600 kg/m^3. This implies that the classic physical picture of inserting a larger hydrophobic molecule into a semi-dense polar solution should involve a larger free energy change than for a smaller molecule might not be entirely accurate. As such these predicted shifts in the solubility trends could have significant industrial applications, since in near supercritical water the effects of corrosion and unwanted side reactions are considerably less severe. In this study we present a more detailed analysis of this predicted behavior by using molecular dynamic computer simulations to calculate the Henry's law constants for two aqueous alkanes at several different temperatures.

4.2 Theory

The solubility of an individual species at infinite dilution in water can be related directly to the Henry's constant ($k_H = \lim_{x_2 \to 0} f_2 / x_2$) through the free energy of hydration (Guillot and Guissani, 1993),

$$k_H = \rho_1 RT \exp[\Delta G^{hyd}(T,P)/RT] \quad , \tag{4.1}$$

where T is the absolute temperature, P is the pressure, R is the ideal gas constant, ρ_1 is the number density of water, x is the mole fraction, f is the fugacity, ΔG^{hyd} is the free energy of hydration according to the definition of Ben-Naim (1992), and the subscripts 1 and 2 refer to the solvent and solute respectively. The free energy of hydration can be described in a rigorous statistical mechanical framework using the relation (Chialvo *et al.*, 1996),

$$\Delta G^{hyd}(T,P) = \int_0^{\rho(P)} \left[\frac{1}{\rho_1} \left(\frac{\partial P}{\partial x_2} \right)_{T,\rho_1}^{\infty} + \frac{\bar{V}_1^0}{\kappa_T} - RT \right] \frac{d\rho_1}{\rho_1} \tag{4.2}$$

where κ_T is the isothermal compressibility of pure water, $\bar{V}_i^0$ is the partial molar volume of the pure component, and the subscript ∞ denotes infinite dilution. It is possible to re-write the expression $(\partial P / \partial x_2)_{T,\rho}^{\infty}$ in terms of the parameter A_{12}^{∞},

$$A_{12}^{\infty} \equiv \lim_{N_2 \to 0} \left(\frac{\partial (PV/RT)}{\partial N_2} \right)_{T,V,N_1} = \frac{\bar{V}_1^0}{RT\kappa_T} + \frac{1}{RT\rho_1} \left(\frac{\partial P}{\partial x_2} \right)_{T,\rho_1}^{\infty} = 1 - C_{12}^{\infty} = \frac{\bar{V}_2^{\infty}}{RT\kappa_T} \tag{4.3}$$

where C_{12}^{∞} is the dimensionless integral, $4\pi\rho_1 \int c_{12}(r) r^2 dr$, for the infinite dilution solute-solvent direct correlation function. As such Eqn. (4.2) simply becomes,

$$\Delta G^{hyd}(T,P) = RT \int_0^{\rho(P)} \frac{(A_{12}^{\infty} - 1)}{\rho_1} d\rho_1 \tag{4.4}$$

and it has been shown experimentally (O'Connell *et al.*, 1996) and theoretically (Chialvo *et al.*, 1996, Levelt Sengers, 1991) that the parameter A_{12}^{∞} remains finite at the critical point. Eqn. (4.4) however does assume a continuous path to the ideal gas standard state exists and for integrations carried out through the two-phase region Eqn. (4.4) are difficult to perform. One possible method to compensate for this difficulty is the use of a simple empirical correction term (Sedlbauer *et al.*, 2000). In principle, since the quantity ΔG^{hyd} is a state variable one should be able to rewrite Eqn. (4.4) using a more complex path that does not cross into the solvent's two-phase region, although this approach has not been examined to our knowledge. It is also worth while to point out that the equation of Plyasunov *et al.* (2000) can not be correctly employed to calculate ΔG^{hyd} at T < T_c because it would involve integration of Eqn. (4.4) through the two-phase region.

At low densities A_{12}^{∞} can be described by a simple virial expansion (O'Connell *et al.*, 1996; McQuarrie, 1976),

$$A_{12}^{\infty} = 1 + 2B_{12}(T)\rho_1 + 3C_{112}(T)\rho_1^{\,2} + 4D_{1112}(T)\rho_1^{\,3} + \ldots \tag{4.5}$$

where $B_{12}(T)$, $C_{112}(T)$, and $D_{1112}(T)$ refer to the 2nd, 3rd , and 4th cross-virial coefficients between the solvent and the solute. The integration of Eqn. (4.4) can easily be carried out yielding the virial expansion of the free energy of hydration (Joslin *et al.*, 1996; O'Connell *et al.*, 1996),

$$\Delta G^{hyd}(T,P) = RT\left[2B_{12}(T)\rho_1 + \frac{3}{2}C_{112}(T)\rho_1^{\,2} + \frac{4}{3}D_{1112}(T)\rho_1^{\,3} + \ldots\right] \tag{4.6}$$

Typically, the radius of convergence for this series is assumed to occur at densities considerably below the critical density (ρ_c), although the actual upper density limit is not well defined. For the case of a simple binary Lennard-Jones system, a recent study by Joslin *et al.* (1996) has found that the virial equation (truncated after the fourth term) gave a good description of the solubilities up to ~2 ρ_c and suggested that the useful density range of Eqn. (4.5) might be considerably larger than previously believed.

4.2.1 Models

In an effort to confirm the possible non-intuitive behavior in the solubility described in part 2, we have used molecular dynamics simulations to calculate the free energy of hydration for two infinitely dilute alkane molecules (butane and octane) at six reduced temperatures (T_r =0.479, 0.541, 0.685, 0.830, 0.974, 1.031, 1.175) along the reduced near-critical isobar P_r = 1.28. For the alkane solutes we have chosen to employ

the united atom model of Martin and Siepmann [TraPPE] (1998) and for the solvent the fluctuating point-charge water model of Rick *et al.* [TIP4P-FQ] (1994).

The TraPPE alkane potential model consists of a string of Lennard-Jones spheres connected to each other via a fixed bond length of 1.54 Å, plus additional carbon-carbon bond bending and dihedral potentials. The TIP4P-FQ model employs the rigid, 4-site TIP4P molecular geometry and allows the partial charges to fluctuate in response to each individual water molecule's unique local electrostatic environment. The model consists of a single Lennard-Jones interaction site on the oxygen, three Coulombic interactions centered on the two hydrogens, and a massless point charge site. The details of the model and description of the algorithm used to propagate the charge degrees of freedom have been described in part 3 (Rick *et al.*, 1994). The Lorentz-Berthelot (Lorentz, 1881; Berthelot, 1898) combining rules for all Lennard-Jones interactions were used, $\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$ and $\sigma_{ij} = (\sigma_{ii} + \sigma_{jj})/2$, and the nonbonding interaction parameters for both the TraPPE and TIP4P-FQ models are reproduced in Table 4.1.

The TIP4P-FQ model for water was selected for a variety of reasons. The TIP4P-FQ model (or any polarizable model for that matter) gives a better description of the local electrostatic environment of water than traditional non-polarizable models [such as the SPC (Berendsen, 1981), SPC/E (Berensen *et al*, 1987) , TIP4P (Jorgensen *et al.*, 1983), or WK (Watanabe and Klein, 1989) models] and appear to provide a more realistic model of the solvent structure (Chialvo *et al.*, 2000). Another advantage of the TIP4P-FQ model is that it uses the experimental gas-phase (*i.e. TIP4P*) geometry. Models using the unrealistic SPC geometry posses an error of ~5° in the HOH angle and ~0.02 angstroms

Table 4.1: Nonbonded Lennard-Jones Parameters for the TraPPE and TIP4P-FQ models.

i	ε_{ii} (kJ/mol)	σ_{ii} (Å)
CH_3	0.8148	3.75
CH_2	0.3824	3.95
Water (Oxygen)	1.197	3.159

in the O-H bond distance. Because the TIP4P geometry is in better agreement with the experimental gas phase and *ab-initio* geometries (Kim and Jordan, 1994) it should lend itself more easily to future quantum mechanical studies than the SPC geometry. This is because at the quantum mechanical level errors in the geometry of this magnitude are significant and some of the most commonly performed *ab-initio* methods [*i.e.* MPx (Møller and Plesset, 1934; Sæbo and Pulay, 1986)] have been shown not to perform well at non-equilibrium geometries (Handy, 1985). Finally, we note the liquid-vapor coexistence curve has been calculated by several independent studies. The Gibbs-Duhem integration (Kofke, 1993) calculation of Yezdimer and Cummings (1999) produced critical constants that are in good agreement with the latest Gibbs ensemble Monte Carlo (Panagiotopoulos, 1987) calculations of Chen *et al.* (2000). However we expect that Chen *et al.*'s vapor phase densities should be more accurate because the initial state point of Yezdimer and Cummings's Gibbs-Duhem integration was based on an approximate Gibbs ensemble Monte Carlo calculation (Medeiros and Costas, 1997) which may have induced errors in the gas phase density. In this study we have taken the critical points of the TIP4P-FQ model to be T_c = 570 K, ρ_c = 300 kg/m^3 and P_c = 19 MPa. These critical constants are similar to those obtained using the SPC model, but not in as good agreement as with experiment as the SPC/E model.

4.2.2 Methodology

Simulations of both aqueous butane and octane where performed at infinite dilution in the NPT ensemble. The simulations of butane involved 256 TIP4P-FQ molecules plus 1 butane-TraPPE molecule, while simulations of octane used 512 TIP4P-

FQ molecules plus 1 octane-TraPPE molecule. All simulations were carried out using a modified version of the DL_POLY software package (Forester and Smith, 1995). Periodic boundary conditions and Ewald summation techniques where employed to reproduce bulk conditions and account for long-range Coulombic interactions (Allen and Tildesley, 1994; Frenkel and Smit, 1996). The values of the largest reciprocal space vector (k_{max}) and the Ewald convergence parameter, α, were determined by the DL_POLY software in order to obtain an accuracy of 10^{-6} in the Coulombic energy and varied between different system sizes. Typical values for k_{max} ranged from 7 to 8 while values of α ranged from 0.18 $Å^{-1}$ to 0.36 $Å^{-1}$. A cut-off radius for the intermolecular interactions was used and taken to be slightly less than half the simulation box length. The Lennard-Jones' energy and pressure were then corrected for the truncation by assuming the radial distribution functions at distances greater than the cut-off were equal to unity. A Nosé-Hoover thermostat (Nosé, 1991) and Andersen barostat (Andersen, 1980) were used to maintain a constant temperature and pressure. A time step of 0.001 picosecond was used and the equations of motion were integrated using a leap-frog routine (Hockney and Eastwood, 1981).

Each state point was allowed to equilibrate for ~100 ps and the free energies of hydration were calculated using a thermodynamic integration (TI) technique (Frenkel and Smit, 1996). The use of TI was advantageous over a Gibbs Ensemble approach because non-idealities in the vapor phase due to the relativity high pressure of the systems need not be considered. In order to avoid the well-documented singularities (Squire and Hoover, 1969; Mezei and Beveridge,1986) that can arise when a simple linear coupling scheme is used with a Lennard-Jones interaction, the nonlinear coupling scheme

proposed by Beutler *et al.* (1994) was used. Their expression for the Lennard-Jones potential was derived by introducing the coupling parameter, λ, into the Lennard-Jones potential curve in an non-trivial manner and takes the following form,

$$U_{ij}^{LJ} = \lambda 4\varepsilon_{ij}\left(\frac{1}{[\gamma(1-\lambda)^2 + (r_{ij}/\sigma_{ij})^6]^2} - \frac{1}{\gamma(1-\lambda)^2 + (r_{ij}/\sigma_{ij})^6}\right) \tag{4.7}$$

where r_{ij} is the interparticle distance and ε_{ij} and σ_{ij} are the energy and size Lennard-Jones potential parameters. The introduction of the arbitrary positive constant γ (which has been set to 0.5 following Beutler *et al.*'s recommendation) successfully removes the singularity of $(dU/d\lambda)_{\lambda=0}$ by converting the Lennard-Jones potential function into a soft-core potential at small values of λ.

Each alkane molecule was grown into solution in one complete step (all alkane interaction sites at once), instead of the more common piece-wise fashion (Gao *et al.*, 1989; Boresh and Karplus, 1999). Müller and Paul (1994) have shown that a single step TI method can be very efficient when dealing with long chains, although their calculations where only conducted on lattice models and only examined the effect of excluded volume. This method has the advantage of being simpler because no dummy atoms and/or alchemy of molecular species are required. One might also expect this method to produce a lower systematic error because there is no propagation of errors for each chain bead. The one possible disadvantage of this approach is however that runs may suffer from significant time-lag hysteresis (see below).

The free energy of solvation, as defined by Ben-Naim (1992), can be determined using the standard expression of a TI,

$$\Delta G^{hyd}(T,P) = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda,T,P} d\lambda \tag{4.8}$$

where U denotes the total intermolecular interaction energy of the system and only the solute-solvent Lennard-Jones intermolecular interactions are coupled as a function of λ. As such, Eqn. (4.8) describes the free energy difference between the solute in the ideal gas state (λ=0; $U^{LJ}_{TraPPE,TIP4P-FQ} = 0$) and the solute in the solution (λ =1). In all of the simulations conducted the average end-to-end distance of the alkane was always considerably smaller than the Lennard-Jones cut-off distance. This assures that there are no explicit solute-solute intermolecular interactions contained in the simulations and allows us to assume that the simulation conditions are representative of infinite dilution. There is no need for any additional corrections to ΔG^{hyd} due to the chain size, solute size, and/or solute structure (Ben-Naim and Mazo, 1993; Martin *et al.*, 1999) in Eqn. (4.8). Differences between the molecular sizes of the solute and solvent can only come into play when deriving intermolecular combining rules (as will be discussed later). Any additional solute rotational-solvent interaction coupling energy is already included by definition in the value of ΔG^{hyd} through the indicated average, $\langle \ldots \rangle_\lambda$.

Each TI consisted of 11 windows at λ = 1.0, 0.8, 0.6, 0.4, 0.3, 0.25, 0.2, 0.15, 0.10, 0.05 and 0. For state points where $T < T_c$, every window was allowed to equilibrate 30 picoseconds and averages of $dU/d\lambda$ where taken for 70 picoseconds. At state points were $T > T_c$, each window was averaged for 110 picoseconds to ensure accurate sampling. The values of $dU/d\lambda$ were then integrated using the trapezoidal rule, and the results are presented in Figure 4.1 and Table 4.2.

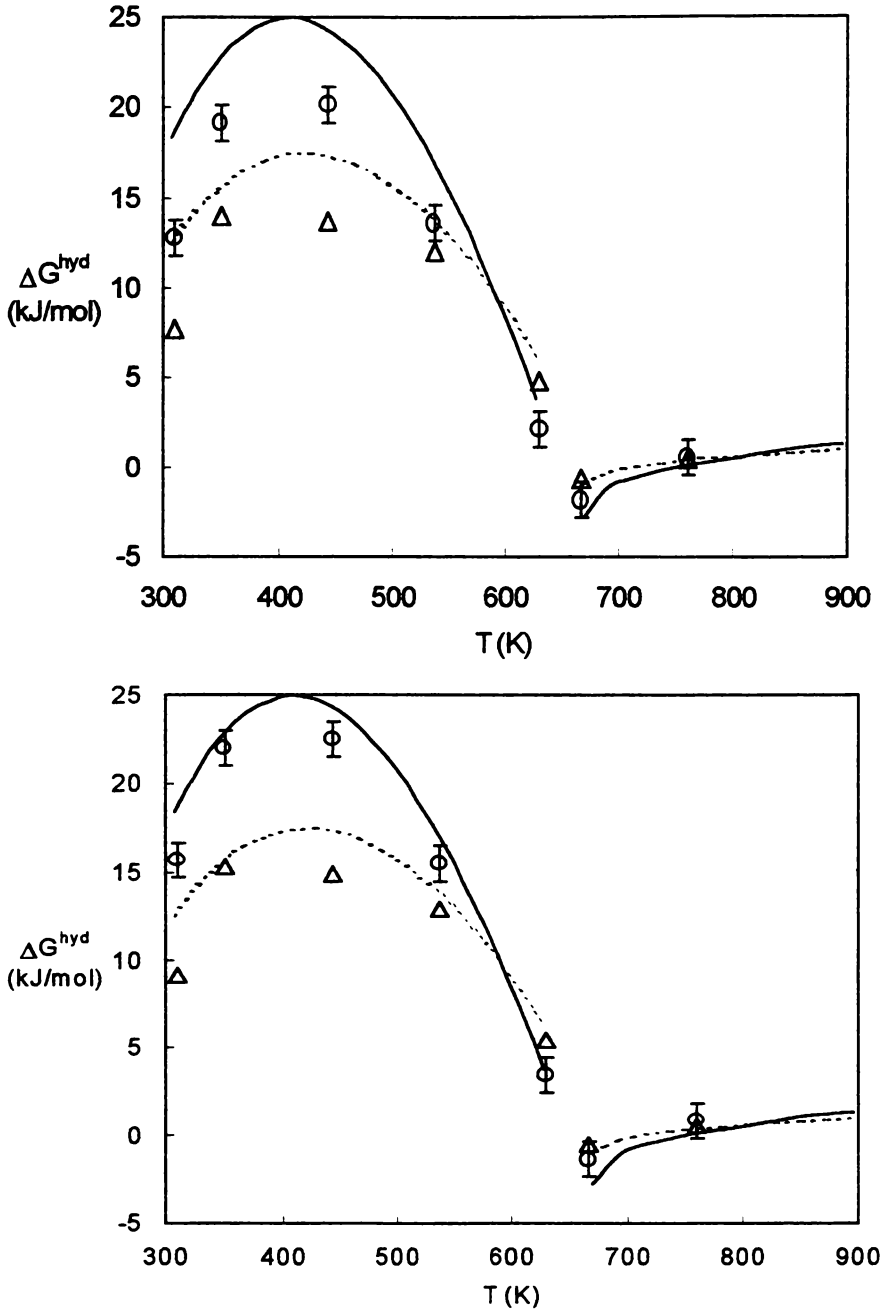

Figure 4.1: ΔG^{hyd} for octane (aq) and butane (aq) as a function of temperature. a) uncorrected b) corrected via Eqn (B.6) OOO- Simulated octane; ΔΔΔ- Simulated butane; ___ - Predicted octane; _ _ _ - Predicted butane (Yezdimer et al. 2000, T < 630 K; Second cross-viral coefficients, T > 660 K). The error in ΔG^{hyd} was found to be on the order of 1 – 1.5 kJ/mol for both octane and butane. For clarity, the error bars are shown only on the octane points. In order to allow a better comparison with the EOS, the x-axis for all of the simulation results have been scaled by a factor of 647 K / 570 K, where 570 K is the critical temperature of the TIP4P-FQ model.

Table 4.2: TraPPE – TIP4P-FQ simulation results.

T_r	ρ_l (kg/mol)[a]	ΔG^{hyd},[b] (kJ/mol)	ΔG_{corr} (kJ/mol)	$\#H_{excluded}$ [c]	B_{12} (cm^3/mol)
		Butane(aq)			
0.478	1013 ± 14	7.67	1.44	11.6	-128
0.541	996 ± 13	14.00	1.31	12.1	-99.4
0.685	906 ± 16	13.72	1.17	9.1	-54.9
0.830	708 ± 24	11.96	0.95	6.2	-26.6
0.974	554 ± 45	4.85	0.62	4.0	-11.2
1.031	129 ± 18	-0.67	0.15	0.9	-5.71
1.175	105 ± 9	0.50	0.13	0.9	6.16
		Octane(aq)			
0.478	1013± 10	12.75	2.93	19.0	-260
0.541	996 ± 10	19.17	2.80	17.4	-195
0.685	908 ± 12	20.15	2.38	13.4	-110
0.830	780 ± 17	13.55	1.90	9.7	-59.6
0.974	565 ± 30	2.14	1.29	6.3	-26.4
1.031	203 ± 25	-1.78	0.50	2.3	-16.4
1.175	106 ± 6	0.59	0.26	1.3	2.08

[a] average solvent density/ errors are the standard deviations

[b] The uncertainty in ΔG^{hyd} was found to be on the order of 1 – 1.5 kJ/mol

[c] The average number of $\#H_{excluded}$ is defined by the number of water hydrogen sites inside the alkane's σ_{CH2} and σ_{CH3} distances

All simulations yielded well-behaved functions for $\langle dU/d\lambda \rangle_\lambda$ and were qualitatively similar to those discussed by Beutler *et al.* A typical example is shown in Figure 4.2. Ideally, a reverse 'growing in' simulation for each alkane and state point should be performed in order to better ascertain the size of any time-lag hysteresis (Lin and Wood, 1996). Unfortunately this would require double the amount of our available computational time. However, the magnitude of this error was estimated by performing a single reverse growing-in simulation for octane at $T_r = 0.541$. At this state point the magnitude of any time-lag hysteresis present in our calculations should be large. It was found that the time-lag hysteresis present at this state point was on the same order as the statistical uncertainty from the simulated values of $\langle dU/d\lambda \rangle_\lambda$, approximately 1.0 to 1.5 kJ/mol.

4.3 Simulated Free Energy Results

The calculated values of ΔG^{hyd} from our simulations show good qualitative agreement with the EOS predictions of Yezdimer *et al.*, and they are able to reproduce the predicted reversal in butane and octane's solvation free energy curves at ~ 620 K. It is tempting to simply connect these shifts in solubility with the increased solvent compressibility near the solvent's critical point (T_c=647 K, P_c=22 MPa); however this possible effect has not yet been rigorously quantified and Chialvo and Cummings (1996, 1994) have shown that the solvation free energy should be independent of any long-range

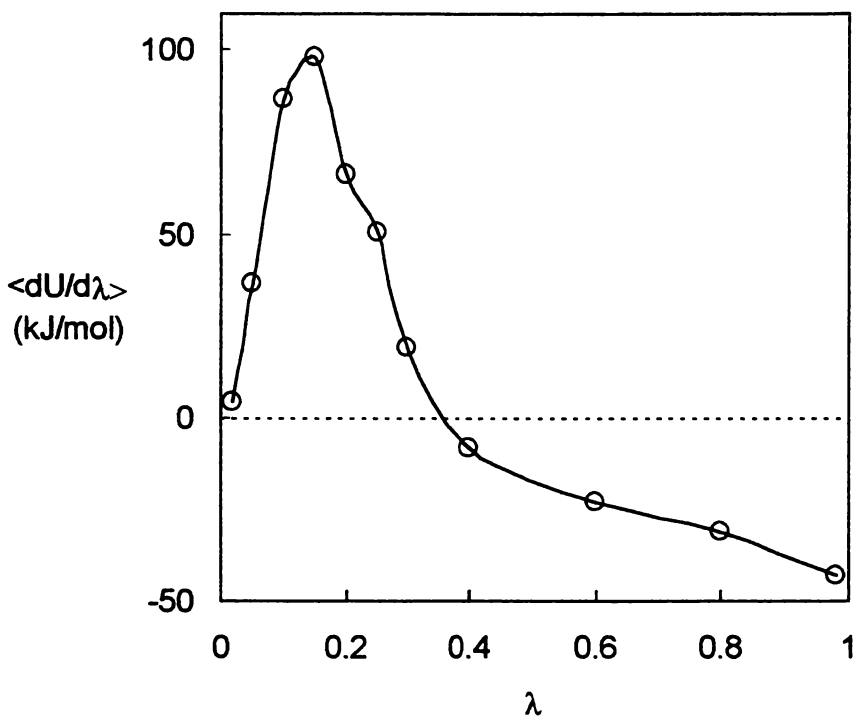

Figure 4.2: Example of $\langle dU/d\lambda \rangle$ for octane (aq) at T_r=0.974, P_r= 1.28

compressibility driven effect. The free energy of hydration calculated by our simulations appears to be finite and continuous at the critical temperature. Interestingly the ΔG^{hyd} curves for both octane and butane appear to show a change from a negative to positive curvature in the supercritical region around 670 K, suggesting that the heat capacity of hydration passes through zero at that point [$\Delta C_p^{hyd} = -T(\partial^2 \Delta G^{hyd} / \partial T^2)_P$].

At supercritical conditions and low densities, the sign and magnitude of ΔG^{hyd} for a given species will be primarily determined by the temperature dependence of $B_{12}(T)$ in Eqn. (4.6). One method of evaluating the leading term in Eqn. (4.6) would be the use of the second-virial correlations developed by Tsonopoulos (Tsonopoulos and Heidmann, 1990; Tsonopoulos, 1974). However, because no experimental data is available at supercritical conditions and alkanes tend to become thermally unstable above 650 K, we have estimated the cross coefficient $B_{12}(T)$ for the TraPPE- TIP4P-FQ models (Figure 4.3). The predicted Boyle temperature (T_b) for the TraPPE- TIP4P-FQ cross virial coefficient $B_{12}(T)$ has been estimated to be ~625 K and ~660 K for butane and octane respectively. However because the critical temperature of our simulations occurs at 570 K rather than the experimental value of 647 K, the T_b for both the butane and octane are both well inside the supercritical region.

At temperatures below T_b, octane is expected to have a larger, negative deviation from ideal gas behavior ($B_{12}(T)<0$) than butane because of the increased number of alkane-water interaction sites. At temperatures above T_b, octane is expected to have a larger positive deviation from ideal gas behavior ($B_{12}(T)>0$) than butane because of octane's larger molecular volume. Because the T_b for both butane and octane occur

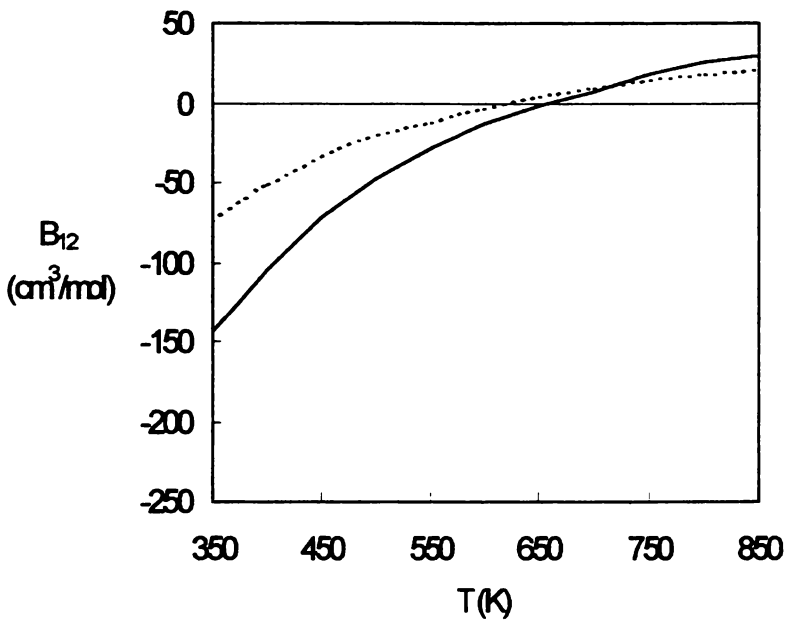

Figure 4.3: Calculated values of $B_{12}(T)$ using the TraPPE and TIP4P-FQ models. ___ - octane(aq); _ _ _ - butane(aq). The values were calculated assuming both butane and octane were in a purely trans configuration. Preliminary calculations of $B_{12}(T)$ showed that differences between the trans and cis conformations was small, less than 5 cm^3/mol.

significantly above the solvent's critical temperature, octane should have a lower free energy of solvation than butane in the region $T_c<T<T_b$, provided that the solvent density is low enough.

Non-negligible contributions from $C_{112}(T)$ and possibly other higher ordered terms are mostly likely contained in our simulated supercritical values of ΔG^{hyd} (T_r=1.031 and 1.175) because the corresponding solvent density (along the P_r=1.28 isobar) is still somewhat high. The presence of these additional contributions to ΔG^{hyd} from $C_{112}(T)$ is also expected from the estimated number of water molecules present near or within the alkane's excluded volume (see below). This may help to explain why the simulation results at T_r=1.031 are somewhat more positive than predicted from only the leading term in Eqn. (4.6). The simulated results for both octane and butane (at T_r=*1.175*), however do begin to show the correct high temperature limit trend (Plyasunov *et al.*, 2000), with octane having a slightly larger free energy of hydration then butane.

The recent equation of state proposed by Plyasunov *et al.* was also examined. It was found that this EOS did not show the reversal in ΔG^{hyd} curves, in either the subcritical or supercritical regions. This result is not surprising because the calculated Boyle temperature for alkane-water interactions from the Plyasunov-O'Connell-Wood equation of state is slightly less than the critical temperature of pure water.

In the high-density region ($T<<T_c$) and along the isobar P_r=*1.28*, higher order correlations between the solute and larger solvent clusters must be considered. These correlations are undoubtedly responsible for the appearance of the so-called hydrophobic and solvent cage effects that occur at lower temperatures. Unfortunately, these contributions can currently only be evaluated in an empirical fashion making a direct

explanation of the initial reversal occurring at T~620 K in terms of a simple virial expansion not feasible. However it is most likely that the reversal in the free energy of hydration occurring around T=620 K is directly related to the solvent's sharp decrease in density in the vicinity of the critical point. The decrease in density serves to decrease the contribution to the free energy of hydration from higher ordered solute-solvent correlations (responsible for solvent cage effects). Thus as the temperature approaches the critical temperature, ΔG^{hyd} begins to move toward the low-density trends.

Figure 4.4 shows a comparison between our simulated values of the Henry's constant (using Eqn.(4.1)) and the EOS predictions. The Henry's constants for aqueous methane and ethane up to 570 K and 65 bars have also recently been calculated by Errington *et al.* (1998). Their calculations were preformed using the Widom test particle insertion technique (Widom, 1963) and employed the TraPPE alkane and SPC/E and MSPC/E (Boulougouris *et al.*, 1998) water models. Good agreement between their results and the experimental values were found (even at low temperatures) and the models were able to accurately predict the maximum in the temperature of dependence of Henry's constant.

It is also important to point out that the sub-critical maximums in the Henry's constant and free energy of hydration occur at different temperatures. The free energy of hydration only measures the amount of free energy required to insert the solute molecule from the ideal gas phase into solution at a particular temperature and pressure. The Henry's constant is related to the inverse of the solubility and must also take into account the amount of available thermal energy (note the additional factor of RT in Eqn. (4.1)).

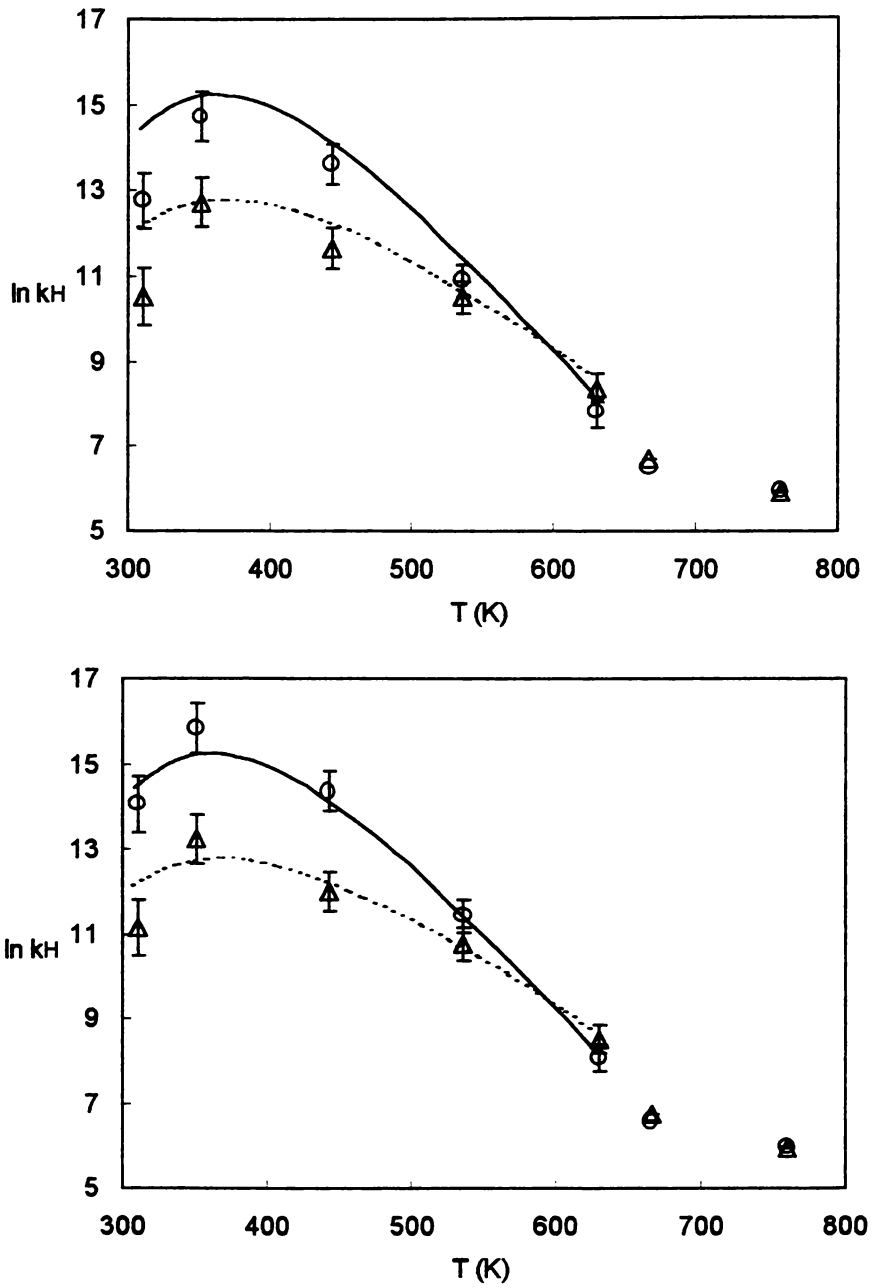

Figure 4.4: Comparison between the EOS and the MD results for $ln\, k_H$ [Eqn. (4.1)]. k_H units in bar. Symbols are the same as in Figure 4.1. In order to allow a better comparison with the EOS, the x-axis for all of the simulation results have been scaled by a factor of 647 K / 570 K, where 570 K is the critical temperature of the TIP4P-FQ model. a) Uncorrected results b) Results corrected using Eqn. B.6

Thus the maximum in the Henry's constant will be weighted toward lower temperatures. As such, the maximum predicted by the EOS in the free energy of hydration (at P_r=1.28) occurs around 425 K for butane and octane while the maximum in the Henry's constant occurs at about 370 K.

Since we are dealing with a mixture, it is important to examine the effect of different combining rules on the calculated values of ΔG^{hyd}. The Berthelot combining rule makes use of the geometric mean assumption, and consequently, the rule is only valid for spherical molecules possessing the same ionization potential and molecular radius. As discussed by Hildebrand and Scott (1962) and by Hirschfelder *et al.* (1954) it is possible to write more precise expressions for the Lennard-Jones well depth combining rules (see Appendix B).

The reason behind the lack of more accurate combining rules for molecular simulation purposes is not clear since they typically represent no additional computational effort. By ignoring dissimilarities in ionization energies, it becomes quite trivial to estimate the error introduced into the total solute-solvent interaction energy due to dissimilarities in the size parameter for the unlike interactions (σ_{12}) for the TIP4P-FQ and TraPPE models [from Eqn.(B.6)]; errors of 2.2%, 3.7%, and 0.2 % in ε_{12} can be expected for the CH_3-TIP4P-FQ, CH_2-TIP4P-FQ, and CH_3-CH_2 interactions, respectively. Although small, these errors can still lead to a significant overestimation of the strength of solute-solvent interactions, especially in the high-density region. Chialvo (1991) has presented a simple method for calculating the errors in ΔG^{hyd} for species whose unlike interaction parameter exhibit small deviations from the Berthelot

combining rule. Another method for estimating the error in ΔG^{hyd} induced by dissimilarities in the size parameter for the united atom and water Lennard-Jones spheres, is the ABC-FEP method of Wood *et al.* (1999). Assuming the ionization energies (defined in the Appendix B) were equal, and by employing Eqn. (B.6) as the perturbing Hamiltonian (instead of the quantum *ab-initio* energies that are normally used in the ABC-FEP method), the error in ΔG^{hyd} can be calculated without additional simulation. The results are presented in Table 4.2, and the corrected values of ΔG^{hyd} are shown in Figure 4.1b. The corrections are largest at the lowest temperatures because of the increased density and average intermolecular interaction energies; at higher temperatures/lower densities the correction becomes small. While the corrected free energies happen to be in general better agreement with the predicted EOS values, it is important to note that significant changes in ΔG^{hyd} can be expected when different combining rules are employed. These facts should be kept in mind when evaluating the accuracy of various potential forcefields.

It is instructive to examine the solvent's relative orientation with respect to the solute molecule. Explicit atom models for alkanes typically posses a dipole moment of around 0.15 D along the C-H bonds. Therefore, there is no energetic orientation preference of the water molecules surrounding the alkane chain in the united atom model. Thus, the water molecules in the first solvation shell of the united-atom alkane can adopt an orientation where the water's hydrogen sites would be pointing inward toward the hydrocarbon's C-H dipole moments. Because of the Coulombic repulsion that would be present if C-H dipole moments were added to configurations generated by the united-

atom model, our simulations would be expected to under predict the amount of orientation disruption of the liquid water structure. To obtain a preliminary quantitative assessment of this possible effect we have calculated the average number of water hydrogens located inside the alkane's excluded volume (see Table 4.2). At the lowest temperature studied it appears that this effect could be significant, with over 11 and 19 hydrogens on average could be existing in potentially unfavorable configurations for butane(aq) and octane(aq), respectively. However the exact magnitude of the free energy change due to these perturbations in solute-solvent orientation is still unknown.

4.4 Summary

In this part we have illustrated how molecular dynamics simulations of infinitely dilute alkane-water mixtures can qualitatively and quantitatively reproduce the reversal in the Henry's constant predicted by Yezdimer *et al.* (2000). The simulation study suggests that the predictions of Yezdimer *et al.* may represent a real physical phenomenon, and may not simply be an artifact of the least squares fitting procedure. The re-reversal of ΔG^{hyd} in the supercritical region can be interpreted in terms of the relative values of the solute-solvent Boyle temperature and solvent critical temperature. The reverse trends in the Henry's constant for increasing alkane chain lengths in the near supercritical region is suggested to be caused by the diminishing of hydrogen bonding structure with increasing isobaric temperature (due to the rapid decrease in solvent density). Although it is unclear to what maximal solute chain length this behavior would persist too, rough predictions of other classes of solutes that might produce a similar reversal in solubility with increasing molecular size or chain length can be made by examining the differences between the

second virial cross coefficient's Boyle temperature and the solvent's critical temperature (*i.e.* acyclic alkenes in water typically have estimated T_b's on the order of 750 K). Examination of other water based solvents with critical temperatures lower than that of pure water may yield a similar organic solubility trends.

Accurate knowledge of the temperature and pressure dependences for organic solubilities is of première importance to the development of new industrial processes. The prediction and understanding of the physics behind the non-intuitive solubility behavior described in this article could possibly lead to new and/or more environmentally friendly industrial organic synthesis and/or separation techniques.

References

Allen, M.P. and Tildesley, D.J. 1994. *Computer Simulations of Liquids,* Oxford Science Publications, Oxford.

Amend, J.P. and Helgeson, H.C. 1997. *Geochim. Cosmochim. Acta,* **61**, 11.

Andersen, H.C. 1980. *J. Chem. Phys.*, **72**, 2384.

Ben-Naim, A. 1992. *Statistical thermodynamics for chemists and biochemists,* Plenum Press, New York.

Ben-Naim, A. and Mazo, R.M. 1993. *J. Phys. Chem.*, **97**, 10829.

Berthelot, D.C.R. 1898. *Hebd. Seanc Acad Sci*, Paris, **126**, 1703.

Berendsen, H. J. C.; Postma, J.P.M.; von Gunsteren, W.F. and Hermans, J. 1981, *Intermolecular Forces*, ed. Pullman, B. (Reidel, Dordrecht), 331.

Berendsen, H.J.C.; Grigera, J.R. and Straatsma, T.P. 1987. *J.Phys. Chem.*, **91**, 6269.

Beutler, T.C.; Mark, A.E.; van Schaik, E.C.; Gerber, P.R.and van Gunsteren, W.F. 1994. *Chem. Phys. Lett.*, **222**, 529.

Boresh, S. and Karplus, M. 1999. *J. Phys. Chem. A.*, **103**, 103.

Boulougouris, G.C.; Economou, I.G. and Theodorou, D.N. 1998. *J. Phys. Chem. B.*, **102**, 1029.

Brock, E.E.; Oshima, Y.; Savage, P.E. and Baker, J.R. 1996. *J. Phys. Chem.*, **100**, 15834.

Chen, B.; Potoff, J.J. and Siepmann, J.I. 2000. *J. Phys. Chem. B*, **104**, 2378.

Chialvo, A.A. 1991. *Mol. Phys.*, **73**, 127.

Chialvo, A.A. and Cummings, P.T. 1994. *AIChE J.*, **40**, 1558.

Chialvo, A.A and Cummings, P.T. 1999. *Avd. Chem. Phys.*, **109**,115.

Chialvo, A.A.; Kalyuzhnyi, Y.V. and Cummings, P.T. 1996. *AIChE J.*, **42**, 571.

Chialvo, A.A.; Yezdimer, E.M.; Driesner, T.; Cummings, P.T. and Simonson, J.M.

2000. *Chem. Phys.*, **258**, 109.

Errington, J.R.; Boulougouris, G.C.; Economou, I.G.; Panagiotopoulos, A.Z. and Teodorou, D.N. 1998. *J. Phys. Chem. B.*, **102** ,8865.

Frenkel, D. and Smit, B. 1996. *Understanding Molecular Simulation*, Acamedic Press, Boston.

Forester, T.R. and Smith,W. 1995. *DL_POLY_2.0* Daresbury Laboratory, Daresbury: Warrington, England.

Gao, J.; Kuczera, K.; Tidor, B. and Karplus, M. 1989. *Science*, **244**, 1069.

Guillot, B. and Guissani, Y. 1993. *Mol. Phys.*, **79**, 53.

Handy, N.C. 1985. *Theor. Chim. Acta*, **68**, 87.

Hildebrand, J.H. and Scott, R.L. 1962. *Regular Solutions*, Prentice Hall In, Englewood Cliffs, New Jersey.

Hirschfelder, J.O.; Curtiss, C.F. and Bird, R.B. 1954. *Molecular theory of Gases and Liquids*, John Wiley & Sons, New York.

Hockney, R.W. and Eastwood, J.W. 1981. *Computer Simulations Using Particles* , McGraw-Hill: New York.

Kim, K. and Jordan, K.D. 1994. *J. Phys. Chem.*, **98**, 10089.

Kirkwood, J.G. and Buff, F.P. 1951. *J. Chem. Phys.*, **19**, 774.

Kofke, D.A. 1993. *J. Chem. Phys.*, **98**, 4149.

Jorgensen, W.L. ; Chandrasekhar, J.; Madura, J.D.; Impey, R.W. and Klein, M.L. 1983. *J. Chem. Phys.*, **79**, 926.

Joslin, C.G.; Gray, G.C.; Goldman, S.; Tomberli, B. and Li,W. 1996. *Mol. Phys.*, **89**, 489.

Levelt Sengers, J.M.H. 1991. *Supercrit. Fluids*, **4**, 215.

Lin, C. and Wood, R.H. 1996. *J. Chem. Chem.*, **100**, 16399.

Lorentz, H. 1881. *Ann. Phys.*, **12**, 127.

Martin, M.G. and Siepmann, J.I. 1998. *J. Phys. Chem. B.*, **102**, 2569.

Martin, M. G.; Zhuravlev, N.D.; Chen, B.; Carr, P.W. and Siepmann, J. I. 1999. *J. Phys. Chem. B*, **103**, 2977.

McQuarrie, D.A. 1976. *Statistical Mechanics*, Harpers Collins Publishers, New York.

Medeiros, M. and Costas, M.E. 1997. *J. Chem. Phys.*, **107** , 2012.

Mezei, M. and Beveridge, D.L. 1986. *Ann. NY, Acad. Sci.*, **482**, 1.

Mitton, D.B.; Orzalli, J.C. and Latanision, R.M. 1995. *Physical Chemistry of Aqueous Systems: Meeting the Needs of Industry* (White, H.J. ; Sengers, J.V. ; Neumann, D.B.; and Bellows, J.C. eds.) Begell House, New York, 638.

Møller, C. and Plesset, M.S. 1934. *Phys. Rev.*, **46**, 618.

Müller, M. and Paul, W. 1994. *J. Chem. Phys.*, **100**, 719.

Nosé, S. 1991. *Progr. Theoret. Phys. Suppl.*, **103**, 1.

O'Connell, J.P. 1994. *Supercritical Fluids: Fundamentals for Application*, NATO ASI Series, v273, eds. E. Kiran and J.M.H. Levelt Sengers, Dordrecht, Holland, 191.

O'Connell, J.P.; Sharygin, A.V. and Wood, R.H. 1996. *Ind. Eng. Chem. Res.*, **35**, 2808.

Panagiotopoulos, A.Z. 1987. *Mol. Phys.*, **61**, 813.

Plyasunov, A.V.; O'Connell, J.P. and Wood, R.H. 2000. *Geochim. et Cosmochim. Acta*, **64**, 495.

Rick, S.W.; Stuart, S.J. and Berne, B.J. 1994. *J. Chem. Phys.*, **101**, 6141.

Sæbo, S. and Pulay, P. 1986. *Theor. Chim. Acta*, 69, 357.

Savage, P.; Gopalan, S. ; Mizan, T.I. ; Martino, C.J. and Brock,E.E. 1995. *AIChe J.*, **41**, 1723.

Sedlbauer, J.; O'Connell, J.P and Wood, R.H. 2000. *Chem. Geol.*, **163**, 43.

Shock, E.L. and Helgeson, H.C. 1990. *Geochim. Cosmochim. Acta*, **54**, 915.

Squire, D.R. and Hoover, W.G. 1969. *J. Chem. Phys.*, **50**, 701.

Tsonopoulos, C. and Heidmann, J.L. 1990. *Fluid Phase Equil*, **57**, 261.

Tsonopoulos, C. 1974. *AIChe J.*, **20**, 263.

Watanabe, K. and Klein, M.L. 1989. *Chem. Phys.*, **131**, 157.

Widom, B. J. 1963. *Chem Phys.*, **39** , 2808.

Wood, R.H.; Yezdimer, E.M.; Sakane, S.; Barriocanal, J.A. and Doren, D.J. **1999**. *J. Chem. Phys.*, **110**, 1329.

Yezdimer, E.M. and Cummings, P.T. 1999. *Mol. Phys.*, **97**, 993.

Yezdimer, E.M.; Sedlbauer, J. and Wood, R.H. 2000. *Chem. Geol.*, **164** , 259.

Yu, J. and Savage, P.E., 1998. *Ind. Eng. Chem. Res.*, **37**, 2.

PART 5

CALCULATING FREE ENERGIES OF SOLVATION USING QUANTUM MECHANICAL INTERACTIONS

5.1 Brief Review of Bulk *Ab-initio* Methods

Methods for calculating free energies of solvation from classical Hamiltonian models are now well developed. These methods have been used to elucidate chemical equilibria in contexts ranging from biochemical to geochemical and have played an important role in increasing our understanding of chemical and phase equilibria. Typical calculations apply molecular dynamics or Monte Carlo simulations to empirically-based classical models with fairly simple analytic descriptions of potential energy as a function of configuration (*e.g.*, Lennard-Jones plus coulombic interactions). Calculating the free energy of solvation for a classical model of solute and solvent is now routine on a modern workstation. The problem with these calculations is that the models are difficult to develop and their accuracy is not easily assessed. As a notable example, there are many models of water–water interactions (Berendsen *et al*, 1987; Watanabe and Klein, 1989; Rick *et al.*, 1994; Chialvo and Cummings, 1996), and yet it is not clear which of these models gives the best overall description of real water. Given the difficulty of developing model potentials, it would be advantageous to have a method that did not require accurate models for every new system explored.

In principle, the model interactions between molecules can be replaced by an *ab-initio* quantum mechanical description of the underlying electronic degrees of freedom. Experience has shown that first-principles calculations predict intra- and inter-molecular interactions with reasonable accuracy in a wide variety of systems, using a common set of approximations. However, in typical applications of these methods, such as energy minimizations, only a small number of configurations are sampled, relative to the number required for a reliable thermal average in a liquid. A direct simulation of a condensed

phase system, with forces calculated from quantum mechanics, is inherently much more expensive than a simulation with classical forces. At present, direct quantum simulations can only be done for small systems on very powerful computers (Laasonen and Klein, 1996; Sprik *et al.*, 1996). Moreover, such methods devote most of the computational effort to generating a set of independent configurations, a task for which accurate energy calculations are probably not needed at every integration temp step. Another approach that is more computationally affordable uses a semiempirical description of the electronic degrees of freedom in a hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) description of solvation (Field *et al.*, 1990; Gao, 1992). Such hybrid calculations are not dramatically more demanding than purely classical simulations, but current semi-empirical methods have a limited range of accuracy (*e.g.*, they typically give a poor description of transition states). A very different approach uses an accurate *ab initio* description of the solute, but treats the solvent as a continuum (Tomasi and Persico, 1994; Tannor *et al.*, 1994). This method neglects specific interactions with solvent, and requires knowledge of the cavity formation energy and the solvent dielectric properties. Calculating the cavity formation energy is always problematic, and in extreme environments, such as a solvent above its critical point, even the dielectric constant may not be well known.

5.2 The ABC-FEP Method

In this part, we describe a method for incorporating quantum mechanical solute-solvent interactions in a calculation of solvation free energies at a much lower cost than for a direct quantum simulation. The critical step is to calculate the difference in free

energy between a solute molecule that interacts with the solvent by a classical energy model, and a solute molecule that interacts with the solvent by a quantum mechanical energy model. This leads to efficient calculation of the free energies of solvation in a hybrid model where solute-solvent interaction energies are calculated from first principles, and solvent-solvent interactions are taken from classical models. The method is less expensive than a full quantum simulation since it uses a classical simulation to generate configurations for thermal averages, and only calculates the quantum interactions at a set of uncorrelated configurations. While the classical description of the solvent-solvent interactions is retained in this model, it is much less demanding to develop new classical models for a few bulk solvents rather for every solute/solvent combination of interest.

We will illustrate the method for a solute, X, in a solvent, S. In the following we will use superscript Q for quantum mechanical energy calculations, C for classical calculations, and a classical Hamiltonian is denoted by H^a. The overall process is shown schematically in Figure 5.1. The first step is to calculate the Gibbs free energy of solvation for a simple classical solute in a classical solvent, $\Delta G[\bullet \rightarrow H^a]$, where $\bullet$ denotes the interaction of a point mass with a solvent (no solute-solvent interactions), and we are using the Ben-Naim and Marcus definition of solvation (Ben-Naim and Marcus, 1984; Lin and Wood, 1996). This may be done by the usual thermodynamic integration (TI) or free energy perturbation (FEP) techniques (Frenkel and Smit, 1996; Allen and Tildesley, 1987). For example, in the FEP approach, one defines a classical Hamiltonian with a parameter, λ, that interpolates between the extremes of a Hamiltonian of a solvent,

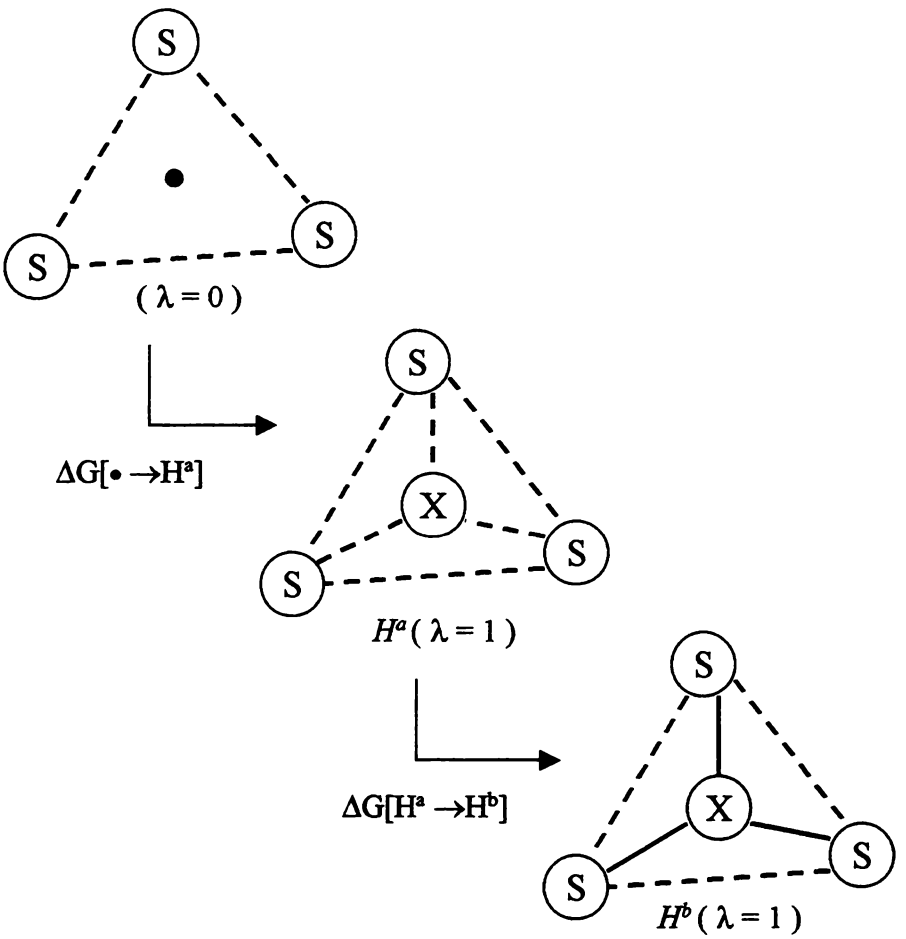

Figure 5.1 : Schematic illustration of the present method for the efficient free energy calculation. The dashed lines (---) indicate classical interactions and solid lines indicate quantum interactions. For clarity the solute, X, is only shown interacting with three solvent molecules, S. The initial configuration represents a point mass solute in the solvent with λ=0. The middle configuration represents the classical solute in classical solvent with $H^a\{S+X\}(\lambda=1)$. The final configuration represents the solute X in solvent S with the hybrid quantum classical interactions and $H^b\{S+X\}$ (λ =1).

S, with non-interacting point mass ($\lambda = 0$) and a Hamiltonian, S+X, with full solute-solvent interactions ($\lambda = 1$). The net change in free energy is determined by choosing several intermediate points, $0 \geq \lambda_i \geq 1$, and summing the incremental changes

$$\Delta G[\bullet \rightarrow H^a] = \sum_i -kT \ln\left\langle \exp[-(H^a(\lambda_{i+1}) - H^a(\lambda_i))/kT \right\rangle_{\lambda_i}$$

$$= \sum_i -kT \ln\left\langle \exp[-(U^C(\lambda_{i+1}) - U^C(\lambda_i))/kT \right\rangle_{\lambda_i} \qquad (5.1)$$

where in the second line it is assumed the masses of the particles are independent of λ, $U^C(\lambda)$ is the total potential energy of the classical model, and the average, $\langle \ldots \rangle_{\lambda_i}$, is calculated from representative configurations of a classical simulation at constant pressure with the Hamiltonian corresponding to $\lambda = \lambda_i$.

The second step in this method (Figure 5.1) is a FEP calculation of $\Delta G[H^a \rightarrow H^b]$, the change in free energy of transforming the classical model solute-solvent interactions into quantum solvent-solute interactions. The total Hamiltonian for the classical simulations can be written as

$$H^a\{S+X\} = U^C_{XS} + U^C_{SS} + U^C_S + U^C_X + T^C \ , \qquad (5.2)$$

where U^C_{XS} indicates the classical interaction energy of solute X with the solvent S, U^C_{SS} indicates the sum of interactions of all solvent molecules with each other, U^C_S indicates the sum of the energies for each single solvent molecule in a vacuum, U^C_X indicates the energy of a solute molecule in a vacuum, and T^C is the classical nuclear kinetic energy. Note that T^C is the same function of momenta throughout the perturbation. For most

classical models $U_S^C = U_X^C = 0$ by definition. We now define a hybrid quantum-classical Hamiltonian by

$$H^b\{S+X\} = U_{XS}^Q + U_{SS}^C + U_S^C + U_X^C + T^C \quad . \tag{5.3}$$

This hybrid model differs from the classical model only in the solute-solvent interactions, which are calculated from a classical potential model in one case and from a quantum mechanical model in the other. We describe how to calculate the necessary components of these interaction energies below.

Using the FEP method, the difference in free energy for systems with Hamiltonians $H^a\{S+X\}$ and $H^b\{S+X\}$ is calculated as,

$$\begin{aligned} \Delta G[H^a \to H^b] &= -kT \ln\left\langle \exp[-(H^b\{S+X\} - H^a\{S+X\})/kT] \right\rangle_a \\ &= -kT \ln\left\langle \exp[-(U_{XS}^Q - U_{XS}^C)/kT \right\rangle_a \\ &= -kT \ln\left\langle \exp[-\Delta U_{XS}[H^a \to H^b]/kT] \right\rangle_a \end{aligned} \tag{5.4}$$

where the average $\langle \ldots \rangle_a$ is over uncorrelated configurations chosen from a classical simulation of the $\{S+X\}$ model with the classical potential ($\lambda = 1$). Note that only the difference between the classical and quantum solute-solvent interaction energies calculated at the same configuration, $\Delta U_{XS}[H^a \to H^b] = U_{XS}^Q - U_{XS}^C$, appears in the expression for $\Delta G[H^a \to H^b]$.

By adding $\Delta G[H^a \to H^b]$ to the free energy of solvating the classical solute in the classical solvent, $\Delta G[\bullet \to H^a]$, one obtains the solvation free energy ($\Delta_s G$) of solute X in solvent S with quantum mechanical solute-solvent interactions and classical solvent-solvent interactions. In this FEP approach (there after referred to as ABC-FEP [AB-initio/Classical Free Energy Perturbation]), the classical model serves three

functions: it provides classical estimates of the cavity and interaction terms through the calculation of $\Delta G[\bullet \rightarrow H^a]$, and it provides a sample of configurations to be used in the average of Eqn. (5.4). The main advantage of this procedure over simulations with fully quantum mechanical interactions is that quantum mechanical calculations only have to be done at representative independent configurations, not at every time step.

To implement this approach, one must be able to calculate U_{XS} independently for the classical and quantum Hamiltonian models. In both classical and quantum calculations one can easily find the total potential energy, U and we can calculate U_{XS} through the traditional decomposition of energy into pair and multi-body interactions. For a configuration without solute X, the total energy is the sum of one-body terms and solvent-solvent interactions,

$$U\{S\} = U_S + U_{SS} = \sum_i u_i + \sum_{i<j} u_{i,j} + \sum_{i<j<k} u_{i,j,k} + \sum_{i<j<k<l} u_{i,j,k,l} + \cdots \quad , \tag{5.5}$$

where i, j, k, l will always be indices on the solvent molecules. This energy may be calculated from either quantum mechanical or classical models. Adding a solute X to this collection of solvent molecules adds the energy of the solute and its interactions with the solvent, so that the total energy is,

$$U\{S+X\} = U_S + U_{SS} + U_X + U_{XS} \tag{5.6}$$

where,

$$U_{XS} = \sum_i u_{X,i} + \sum_{i<j} u_{X,i,j} + \sum_{i<j<k} u_{X,i,j,k} + \cdots \tag{5.7}$$

and the index X refers to the solute molecule. Note that while there are many-body contributions to U_{XS} that involve multiple solvent molecules, these interactions are

distinct from those included in U_{SS}. Thus, for a given configuration, the full solvent-solute interaction energy is

$$U_{XS} = U\{S+X\} - U\{S\} - U_X \tag{5.8}$$

In other words, for either quantum mechanical or classical models, U_{XS} can be obtained from a total energy calculation of the configuration containing solvent S plus solute X by subtracting the total energy of the system in exactly the same solvent configuration without the solute particle, and the total energy of an isolated molecule X.

5.3 Variants of the Method

The present method is not limited to solvents with only one type of solvent molecule or to systems with a solute at infinite dilution. For instance, the free energy of solvation of CO_2 in an H_2O-CO_2-NH_3 mixture is easily calculated by growing a classical CO_2 molecule into a classical solvent mixture and transforming this molecule into a particle with quantum interactions with all surrounding "solvent" molecules (including other CO_2 molecules). Repeating this procedure for all components allows the chemical potentials of all the components in a complicated mixture to be calculated. Solvent effects on the activation barriers for chemical reactions can be calculated by this method by comparing the difference in free energy of the reactants at infinite separation and the reactants at any point along the reaction path. Another variant of this method can be used to calculate the free energy of a pure fluid, by first calculating the free energy of a classical model of the fluid and then using the FEP method to find the effect of changing the classical model energies into quantum mechanically-derived energies. Finally, the

"solvent' need not be a fluid, so free energies of transforming a "solute" in a solid or on a surface into a different species can easily be calculated. In each case, all that is necessary is the free energy of transformation for classical models of the different species and $\Delta G[H^a \rightarrow H^b]$ calculated for the transformation of each classical model to a quantum model.

5.4 Implementation of the Method

The quantum mechanical part of the calculation can be carried out in various approximations. The most direct approach is to use the same periodic boundary conditions in both the quantum and classical calculations. Density functional theory (DFT) is the natural method to use in such a calculation, since most other correlated first-principles methods are not easily applied to periodic systems. Current versions of DFT do not reliably model dispersion forces, so this may not be adequate in situations where such forces have a large effect on the energy. An alternate approach, which introduces new approximations, is to apply the quantum mechanical calculations to a cluster excised from the classical configuration. The smallest reasonable cluster would include the first solvation shell, but the reliability of the approximation will be increased if both the first and second solvation shells are included. One advantage of using cluster models is that a wide variety of electronic structure methods can be applied, so that different methods can be compared. While the focus of this part is on first-principles methods as corrections to classical models, the same approach could be applied using semi-empirical methods to calculate the quantum interaction energies, or using a hybrid QM/MM simulation to

generate configurations. In the present model, vibrations are treated classically, but quantum corrections could be added if they were important.

5.5 Test of the method

In order to verify that the present method can give accurate free energies, we have chosen for our test case the calculation of the hydration free energy of various water models as solutes in a polarizable model of water at ambient conditions. We have also tentatively chosen criteria for identifying classical models that provide a good sampling of configurations so that only an affordable number of quantum calculations are needed. We have designed three different tests in order to establish if our new method with those criteria can 1) recognize a *good* classical solute model, 2) give accurate results with a *good* classical solute model with a reasonable number of independent configurations, and 3) identify a *bad* classical solute model and improve the model using values of $\Delta U_{XS}[H^a \rightarrow H^b]$ (so that the method is self-correcting).

For computational efficiency we have adopted a cluster method to calculate $\Delta U_{XS}[H^a \rightarrow H_n^b]$ using the nearest n solvent molecules. The free energy change for growing in a classical solute model and then transforming the interactions with the nearest n solvent molecules to quantum interactions is

$$\Delta_s G = \Delta G[\bullet \rightarrow H^a \rightarrow H_n^b] = \Delta G[\bullet \rightarrow H^a] + \Delta G[H^a \rightarrow H_n^b] \qquad (5.9)$$

where $\Delta G[\bullet \rightarrow H^a]$ is the classical hydration free energy of growing a point mass into solute X using H^a for all of the solute-solvent interaction energies, and $\Delta G[H^a \rightarrow H_n^b]$ is the free energy of transforming Hamiltonian a into Hamiltonian b using Eqn. 5.4 with

only the closest n solvent molecule interactions perturbed. The hybrid quantum-classical Hamiltonian in our cluster model approximation is given by,

$$H_n^b = U_{XS,n}^Q + U_{XS,N-n}^C + U_{SS}^C + U_S^C + U_X^C + T^C \tag{5.10}$$

where N denotes the total number of molecules in the system. The classical solvent model for all calculations has been chosen to be the non-pairwise additive, fluctuating charge model (TIP4P-FQ) of Rick *et al.* (1994). We have investigated six different water models for the solute, namely TIP4P-FQ, TIP4P, M1, M2, M3, and M4 (the four latter models will be discussed in detail below). The parameters of all the models used are given in Table 5.1.

The configurations used in the calculation of $\Delta U_{XS}[H^a \rightarrow H_n^b]$ were generated by a molecular dynamics simulation containing one solute, X, molecule and 127 TIP4P-FQ molecules. The simulations were ran using a modified DL_POLY software package (Forester and Smith, 1995) using the NVT ensemble and the temperature and density were 298.15 K and 0.997 g/cm^3 (pressure $\approx$ 1 bar). Periodic boundary conditions were employed and the longer ranged Coulomb interactions were treated using the Ewald sum technique. The Lagrangian equations of motion (McQuarrie, 1976) were integrated using a time step of 0.5 fs and each system was allowed to equilibrium for over 40 ps . Examination of the auto correlation functions indicated that independent configurations could be sampled at 2 ps intervals. $U\{S+X\}$ and $U\{S\}$ for an isolated cluster containing n solvent molecules from each independent configuration were then calculated using Eqn. 5.5 and Eqn. 5.6. In this calculation, the partial charges on the TIP4P-FQ molecules were adjusted to their minimum energy values for both isolated clusters. This

Table 5.1: Various classical water model potential parameters[a]

Model *i*	ε_{ii} (kJ/mol)	$\varepsilon_{i,FQ}$(kJ/mol)	σ_{ii} (Å)	$\sigma_{i,FQ}$ (Å)	q_H (e)
TIP4P-FQ(FQ)	1.197	---	3.159	---	---
TIP4P	0.6485	0.8811	3.154	3.157	0.52
M1	2.957	1.881	5.048	4.104	0.680
M2	2.024	1.557	2.851	3.005	0.450
M3	0.5445	0.8073	3.323	3.241	0.523
M4	0.4135	0.7035	3.379	3.269	0.528

[a] The Lennard-Jones parameters (ε and σ) of model *i* are given in terms of the *ii* interactions. For solute solvent-interactions we have used the Berthelot combining rules, $\varepsilon_{i,FQ} = \sqrt{\varepsilon_i \varepsilon_{FQ}}$ and $\sigma_{i,FQ} = (\sigma_i + \sigma_{FQ})/2$.

is necessary to insure that $U\{S+X\}$ and $U\{S\}$ contain only interactions between the first *n* molecules.

For the calculations of the classical hydration free energy, $\Delta G[\bullet \rightarrow H^a]$, we have used the thermodynamic integration (TI) method in two steps. Rick and Berne (1996) have shown how to perform TI on a fluctuating charge solute model. First we grow a point mass to a Lennard-Jones particle with zero charges on the atom sites ($\lambda_1 = 0 \rightarrow 1$) then "charge up" to the desired solute potential ($\lambda_2 = 0 \rightarrow 1$). The free energy is calculated by,

$$\Delta G[\bullet \rightarrow H^a] = \int_0^1 \left\langle \frac{\partial H^a}{\partial \lambda_1} \right\rangle_{\lambda_2=0} d\lambda_1 + \int_0^1 \left\langle \frac{\partial H^a}{\partial \lambda_2} \right\rangle_{\lambda_1=1} d\lambda_2 \tag{5.11}$$

Eleven values of λ's equally spaced from 0.0 to 1.0 were used for both λ_1 and λ_2. Both the Lennard-Jones potential and Coulombic potential were grown in and grown out of the solution. The derivatives were approximated by a finite difference slope at each λ using $\Delta\lambda = 0.01$ ($\langle \partial H^a/\partial\lambda \rangle \approx \langle \Delta H^a/\Delta\lambda \rangle$), and the averages were taken over a total of 660 ps for the forward and backward runs of λ_1 and λ_2. The slopes were integrated by a polynomial fit (Hummer and Szabo, 1996; Sakane *et al.*, 1998) (Simpson's rule was used when the curve was not smooth), and estimates of statistical error were determined using a block average.

For convenience our simulations were performed with the NVT ensemble so that we have obtained the Helmholtz free energy of solvation ($\Delta A[\bullet \rightarrow H^a]$). $\Delta A[\bullet \rightarrow H^a]$ is equal to $\Delta G[\bullet \rightarrow H^a]$ if a enough solvent molecules are included in the simulation. For

the simulation with the finite number of solvent molecules, the difference between these free energies can be estimated from the extra work of pushing back the solvent at constant volume $\Delta A[\bullet \rightarrow H^a] - \Delta G[\bullet \rightarrow H^a] \approx \Delta V \Delta P/2$, where ΔV is the (partial) molar volume of the solute, and ΔP is the pressure difference between the states $\lambda = 0$ and 1. In our test case, the estimated $\Delta V \Delta P/2$ was 0.16 kJ/mol for the solvation of a water molecule in 127 solvent water molecules at ambient conditions. Our reported values of $\Delta G[\bullet \rightarrow H^a]$ in this article have been corrected.

All the quantum energy calculations were conducted using the Gaussian94 software package (1994). The basis set was chosen according to two criteria, dimer energy and monomer dipole moment. Earlier studies (Kim and Jordan, 1994) have found that the aug-cc-pVDZ basis set with DFT-B3LYP predicts a dimer energy of –19.7 kJ/mol and a dipole moment of 1.854 Debye compared to the experimental values of –23 ± 3 kJ/mol and 1.854 Debye (Curtis *et al.*, 1979; Dyke and Muenter, 1973). Energies for U_{XS}^{Q} were calculated from the classically generated configurations using Eqn. 5.8. Thus one can easily obtain the solute-solvent interaction energies by calculating the total energy of both {S+X} and {S}.

The interaction energies reported here do not account for relaxation of the water geometry in the solvated and unsolvated environments. Moreover, the DFT calculations have not been corrected for basis set superposition error (BSSE), and the DFT functional is known to underestimate dispersion forces. The BSSE can be estimated by the counterpoise method (Boys and Bernardi, 1970), and calculations on a sample of ten configurations (chosen from the 50 used to calculate configurational averages) show that

the BSSE makes the interaction energy of solute with the solvent in the first shell (n = 5) more negative by 4.9 ± 0.7 kJ/mol. The BSSE is partly compensated by the underestimate of dispersion forces which makes the interaction energy less negative. An estimate of the net error in our DFT interactions can be obtained by comparing pair interaction energies from DFT to those from symmetry adapted perturbation theory (SAPT), a method that accurately predicts interaction energies without BSSE (Mas and Szalewicz, 1997). When averaged over pairs of molecules in the first shell over 95 configurations, DFT pair interactions are less negative than SAPT interactions by 1.2 ± 1 kJ/mol. These errors can accumulate in clusters, so further investigations of the errors and improved first-principles calculations will be necessary to achieve higher accuracy in the free energy calculations.

5.5.1 Sampling Criteria

We investigated four different criteria for identifying whether a classical model gives *good*, *marginal*, or *bad* samples of configurations for calculating $\Delta G[H^a \rightarrow H_n^b]$. If a model is *good*, the energy analysis should show clearly that 1) the average change in energy in Eqn. 5.4, $\left\langle \Delta U_{XS}[H^a \rightarrow H^b]/kT \right\rangle_a \equiv \left\langle \Delta U/kT \right\rangle$, is small, 2) the standard deviation of the energies, $\sigma[\Delta U/kT]$, is small, 3) $\exp(-\Delta U/kT)$ in Eqn. 5.4 is adequately sampled by the independent configurations (N_{conf}), and 4) $\Delta G[H^a \rightarrow H_n^b]$ is small.

In order to see if the classical model has adequately sampled $\exp(-\Delta U/kT)$ we have plotted the histograms of $\Delta U/kT$ and $\exp(-\Delta U/kT)$ (Allen and Tildesley, 1987).

We have divided the configurations into bins and plotted both N_{bin}/N_{conf} versus $\Delta U_{bin}/kT$ and $\sum_{i=1}^{N_{bin}} \exp(-\Delta U_i/kT) \Big/ \sum_{j=1}^{N_{conf}} \exp(-\Delta U_j/kT)$ versus $\Delta U_{bin}/kT$, where N_{bin} is the number of configurations in a bin. By examining the distribution of $\exp(-\Delta U/kT)$ one can see if the peak of the distribution is well sampled by the configurations. If so, $\Delta G[H^a \rightarrow H_n^b]$ should be accurate. Conversely, if the most of the peak of the distribution has not been sampled then Eqn.5.4 will not give accurate results because of the limited number of configurations used, signaling that more configurations or a better classical model are needed. Based on the results to be described below we have tentatively classified models as *good* when $\langle \Delta U/kT \rangle$ and $\sigma[\Delta U/kT]$ are less then 1.5, most of the peak in $\exp(-\Delta U/kT)$ is sampled, and $\Delta G[H^a \rightarrow H_n^b]$ is less than 2 kJ/mol. *Bad* models are those with $\langle \Delta U/kT \rangle$ and $\sigma[\Delta U/kT]$ greater than 2, and poor sampling of the peak in $\exp(-\Delta U/kT)$. *Marginal* models are in between these two extremes. Wood *et al.* (1991) have shown that for a Gaussian distribution of ΔU and N_{conf}=100 the systematic error due to bad sampling of the exponential increases very rapidly for $\sigma > 2\ kT$ (e.g. for $\sigma[\Delta U/kT] = 2$, 4 and 9 the error is 0.16 kT, 2.0 kT, and 23 kT, respectively). Thus $\sigma <$ 1.5 is a reasonable criterion for a *good* model with a Gaussian distribution.

5.5.2 Test of the Models

First we tried our best polarizable classical model for solute X, namely the fluctuating charge model (H^a = TIP4P-FQ), to see if it behaves as a *good* model and

gives accurate results with a reasonable number of independent configurations (N_{conf}). The results of the hydration free energy of water with H^a = TIP4P-FQ and

$$H_n^b = DFT_n = U_{XS,n}^{DFT} + U_{XS,N-n}^{FQ} + U_{SS}^{FQ} + U_S^{FQ} + U_X^{FQ} + T^{FQ} \tag{5.12}$$

are presented in Table 5.2, where DFT_n indicates that the solute interaction energy with the first n solvent molecules is the DFT energy. The hydration free energy calculated for n = 5, 9, and 12 agree with the experimental value of –26.5 kJ/mol (Sanderson, 1951; Ben-Naim and Marcus, 1984) within our estimated statistical error (± 2 kJ/mol). These differences are less than our expectations for the possible errors in DFT calculations, so we conclude that a *good* classical model can give accurate results with even N_{conf} = 25 and n = 5, 9, or 12. Table 5.2 and Figure 5.2-a show that by our classification scheme, FQ is a *good* model for DFT interactions in the first shell (n = 5) and a *marginal* or *bad* model when second shell waters are included (n = 9 or 12). (The larger $\Delta U/kT$ for n = 9 and 12 may be due to underestimation of the dispersion energies in DFT calculations). We presume the TIP4P-FQ model is accurate in the outer shells because the classical to quantum interaction transformations yield good results for the first solvation shell (n = 5) and $\Delta G[\bullet \rightarrow H^a]$ was found to be quite close to the experimental value so the total free energy mimics that of real water.

Next we used our best non-polarizable model, TIP4P, as the solute X model. The results for the first solvation shell (n = 5) show that TIP4P is a *bad* model for DFT water in the first shell (Table 5.2). The histogram shows that the energies were not sampled adequately (Figure 5.2-b), while Table 5.2 shows that $\langle \Delta U_{XS}[TIP4P \rightarrow DFT_5]/kT \rangle$ is too large ($\approx$ 4), $\sigma[\Delta U_{XS}[TIP4P \rightarrow DFT_5]/kT]$ is too large ($\approx$ 1.9) and

Table 5.2: Test of the ABC-FEP method with $H^a = FQ$ and $TIP4P$, $H_n^b = DFT_n$.

n	N_{conf}	$\langle \Delta U/kT \rangle$	$\sigma(\Delta U/kT)$	$\Delta G[H^a \rightarrow DFT_n]^*$ (kJ/mol)	$\Delta G[\bullet \rightarrow H^a \rightarrow DFT_n]$ (kJ/mol)
		H^a = TIP4P-FQ , $\Delta G[\bullet \rightarrow (TIP4P - FQ)]$ = -26.2 ± 2.0 (kJ/mol)			
5	25	-0.2 ± 0.2	0.9	-1.4 ± 0.5	-27.6 ±2.1
	50	0.2 ± 0.2	1.1	-0.7 ± 0.4	-26.9 ± 2.0
	95	0.4 ± 0.1	1.4	-1.0 ± 0.5	-27.2 ± 2.1
9	25	1.4 ± 0.3	1.6	0.5 ± 1.8	-25.7 ± 2.7
	50	2.0 ± 0.2	1.7	1.2 ± 1.3	-25.0 ± 2.4
	95	2.1 ± 0.2	1.7	0.3 ± 1.6	-25.9 ± 2.6
12	25	1.8 ± 0.3	1.6	1.8 ± 1.0	-24.4 ± 2.2
	50	2.4 ± 0.3	1.9	2.2 ± 0.9	-24.0 ± 2.2
		H^a = TIP4P, $\Delta G[\bullet \rightarrow TIP4P]$ = -25.6 ± 1.9 (kJ/mol)			
5	25	4.2 ± 0.4	2.0	6.0 ± 1.8	-19.6 ± 2.6
	50	4.0 ± 0.3	1.9	6.3 ± 0.9	-19.3 ± 2.1
	100	4.0 ± 0.2	1.9	6.1 ± 0.8	-19.5 ± 2.1

* The estimated errors are
$\pm Max\left\{-kT\log[\langle \exp(-\Delta U/kT)\rangle \pm \delta\langle \exp(-\Delta U/kT)\rangle] - [-kT\log\langle \exp(-\Delta U/kT)\rangle]\right\}$
where $\delta\langle \exp(-\Delta U/kT)\rangle$ is the statistical error from $\exp(-\Delta U/kT)$. This estimate does not include possibly large systematic errors when the peak in the exponentially weighted histogram is not well sampled.

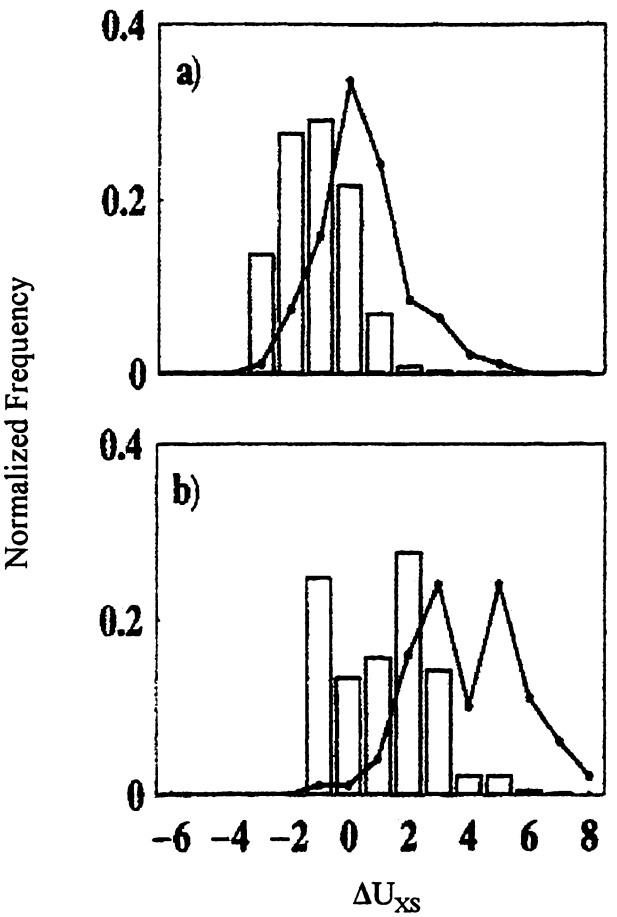

Figure 5.2: Energy distributions of a) ΔU_{XS}[FQ→DFT] and b) ΔU_{XS}[TIP4P→DFT] for N_{conf} = 95. In figure a) most of the peak of the weighted histogram appears to be sampled, where as in figure b) it is not certain that most of the peak has been sampled.

Solid line: N_{bin}/N_{conf}

Histogram: $\sum_{i=1}^{N_{bin}} \exp(-\Delta U_i/kT) \Big/ \sum_{j=1}^{N_{conf}} \exp(-\Delta U_j/kT)$

$\Delta G[TIP4P \rightarrow DFT_5]$ is large ($\approx$ 6 kJ/mol). The classical free energy of growing TIP4P was found to fall close to the experimental value ($\Delta G[\bullet \rightarrow TIP4P]$ = -25.6 ± 1.9 kJ/mol). This indicates that averaged over all solvation shells, TIP4P gives the correct free energy of hydration and that the more attractive interactions in the first shell are compensated by more repulsive interactions in the outer shells. The question of whether DFT or TIP4P is a better model for real water is irrelevant to the present discussion.

5.5.3 Iterative Improvement of Classical Models

Now we wish to see if we can diagnose that a model is *bad* and improve the model using a fitting procedure to minimize the solute-solvent interaction energy differences, ΔU. Since the DFT calculations are time consuming and we already have shown that the TIP4P-FQ model yields a good description of DFT interaction energies in the first shell and the correct total free energy at ambient conditions, we have used TIP4P-FQ water as a replacement for an accurate quantum water model ($H_n^b = TIP4P - FQ_n$), and we have chosen to use N_{conf} = 50 for the remainder of our investigations. Thus only classical calculations are required for this test and model improvements can more quickly be seen. We began our test by creating a very *bad* water model, M1, by increasing $\sigma_{i,FQ}$, $\varepsilon_{i,FQ}$, and q_H of the TIP4P model by 30% (see Table 5.1). When we apply our method to M1 (= H^a) with N_{conf} = 50 and n = 5, we can easily see that M1 is a *bad* model from Figure 5.3-a and Table 5.3. While it is true that $\langle \Delta U / kT \rangle$ is small, this is presumably because the decrease in energy due to the large ε and qH is compensated by the increase in energy due to the large σ of the M1 model in this

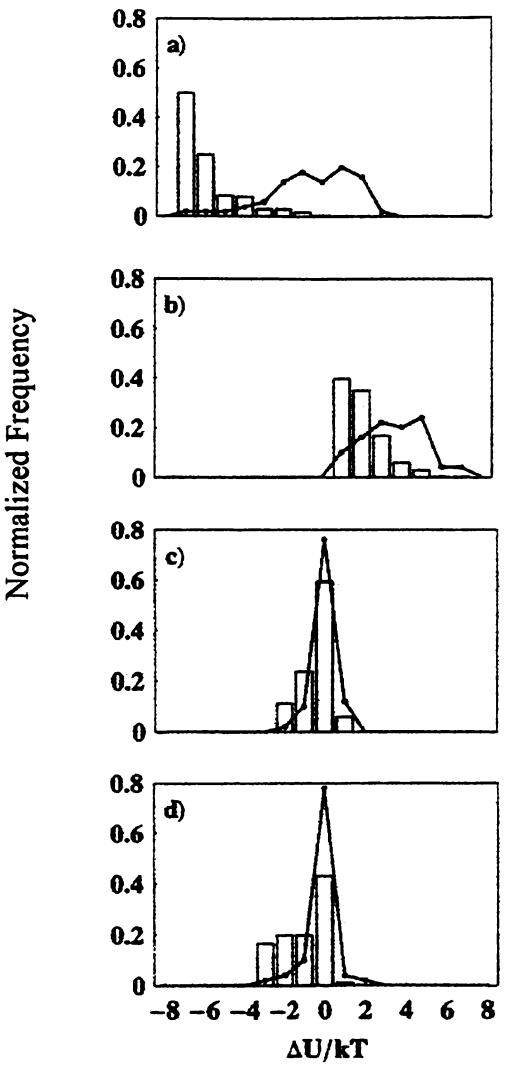

Figure 5.3: Energy distributions of the MX models. a) ΔU_{XS}[M1→DFT] ; b) ΔU_{XS}[M2→DFT] ; c) ΔU_{XS}[M3→DFT] ; d) ΔU_{XS}[M4→DFT] . The bad model, M1, in figure a) has a very large σ and most of the weighted histogram's peak has not been sampled. Model M2 in figure b) over corrects the parameters, <U/kT> is now too positive, *σ[U/kT]* is smaller but still too large, and most of the peak of the weighted histogram is not well sampled. Both models M3 and M4 have low <U/kT>, low *σ[U/kT]* and most of the peak in the histogram is sampled.

Solid line: N_{bin}/N_{conf}

Histogram: $\sum_{i=1}^{N_{bin}} \exp(-\Delta U_i/kT) \Big/ \sum_{j=1}^{N_{conf}} \exp(-\Delta U_j/kT)$

Table 5.3: Recognizing and improving a *bad* model (n = 5, N_{conf}= 50)

H^a	$\left\langle \Delta U_{xs}\left[\begin{matrix}H^a \rightarrow \\ (TIP4P-FQ)_5\end{matrix}\right]/kT \right\rangle$ (kJ/mol)	$\sigma\left\{\Delta U_{xs}\left[\begin{matrix}H^a \rightarrow \\ (TIP4P-FQ)_5\end{matrix}\right]/kT\right\}$ (kJ/mol)	$\Delta G\left[\begin{matrix}H^a \rightarrow \\ (TIP4P-FQ)_5\end{matrix}\right]^*$ (kJ/mol)
M1	-0.6 ± 0.3	2.1	-8.4 ± 2.0
M2	3.6 ± 0.2	1.5	6.7 ± 0.5
M3	0.0 ± 0.1	0.5	-0.4 ± 0.3
M4	0.1 ± 0.1	0.8	-1.2 ± 0.6

* The estimated errors are
$\pm Max\left\{-kT\log[\langle\exp(-\Delta U/kT)\rangle \pm \delta\langle\exp(-\Delta U/kT)\rangle] - [-kT\log\langle\exp(-\Delta U/kT)\rangle]\right\}$
where $\delta\langle\exp(-\Delta U/kT)\rangle$ is the statistical error of $\exp(-\Delta U/kT)$. This estimate does not include possibly large systematic errors when the peak in the exponentially weighted histogram is not well sampled.

particular case (as expected, a small $\langle \Delta U / kT \rangle$ is not a sufficient criteria for a *good* model).

In an attempt to improve the M1 potential, the interaction energy difference ΔU for the 50 first shell (n = 5) configurations was minimized in a least-squares procedure by readjusting the parameters of M1 ($\sigma_{i,FQ}$, $\varepsilon_{i,FQ}$, and q_H). The iterative scheme is shown is Figure 5.4. The set of 50 independent configurations that were generated using the M1 potential as the solute were used to calculate both U_{XS}^{C} and U_{XS}^{Q} with n = 5. For these configurations the parameters of the classical model (M1 with parameters $\varepsilon_{M1,FQ}$, $\sigma_{M1,FQ}$, and q_H in the present case) were varied to make $\sum_i (U_{XS}^{C} - U_{XS}^{Q})^2$ as small as possible. Mezei (1991) used a similar procedure to refine a pairwise additive water model. The best fit $\varepsilon_{M1,FQ}$, $\sigma_{M1,FQ}$, and q_H for M1 were then designated to be the M2 model potential parameters($\varepsilon_{M2,FQ}$, $\sigma_{M2,FQ}$, q_H) . This procedure was repeated twice more, generating models M3 and M4. The results in Table 5.3 and Figure 5.3 show that the first cycle of refinement over-corrected $\sigma_{i,FQ}$, $\varepsilon_{i,FQ}$, and q_H but, after one more cycle, the values of $\sigma_{i,FQ}$, $\varepsilon_{i,FQ}$, and q_H have effectively converged. By our tentative criteria, both M3 and M4 are *good* models for TIP4P-FQ water in the first shell (n = 5). Thus, only two cycles of a simple least-squares fitting procedure, using only 50 configurations in each cycle, was able to transform a *bad* model into a *good* model for TIP4P-FQ water in the first shell. This example indicates that refining a *bad* model to improve configuration sampling can be far more efficient than increasing N_{conf}.

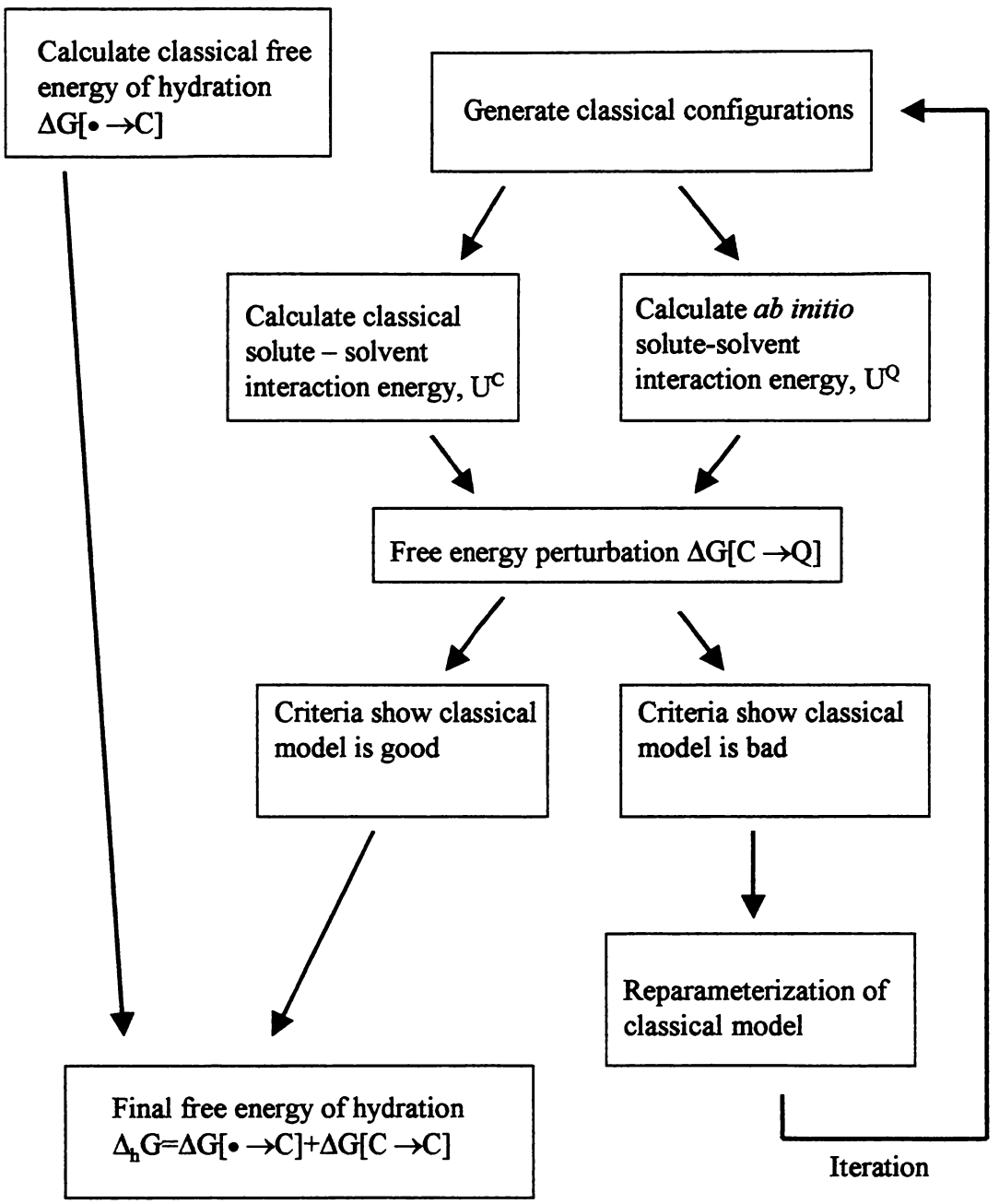

Figure 5.4 : Process diagram for the ABC-FEP method

Note that, in this test case, we have obtained an improved model (M4) that gives correct interaction energies for the first shell, but it is not necessarily a good model for bulk water properties. In full applications of our method, it would be possible to use ΔU 's from larger systems for the least-square fit in order to create a good model for bulk water. However since our interest in the test was in obtaining a model which gives adequate sampling for our FEP method (starting from a bad water model) and was *not* to produce a brand new water model, we have not investigated the bulk properties of M4. In the next section we will show that this defective model (M4) is sufficient to obtain accurate $\Delta_s G$ when large enough system is used for the FEP.

5.5.4 Cluster Size Effects

Finally we examined the accuracy of our method as a function of n for $N_{conf} = 50$, again using the TIP4P-FQ model as a stand-in for the DFT model. Previously we had found that TIP4P is a *bad* model for DFT water in the first shell even though the bulk properties for TIP4P are correct. The M4 model, on the other hand, is a *good* model for TIP4P-FQ water in the first shell but not necessarily for the outer shells. Here we wanted to see if the ABC-FEP method could yield the free energy of hydration predicted by the TIP4P-FQ model by using two different initial classical Hamiltonians, namely the TIP4P and the M4 models. For these two models, the hydration free energies were calculated using several cluster sizes ($0 \le n \le 34$, where the value for $n = 0$ indicates no transformation from H^a to TIP4P-FQ) (Figure 5.5 and Table 5.4). The hydration free energy of M4, before any transformation to the TIP4P-FQ model ($n = 0$), is -17.0 ± 1.6

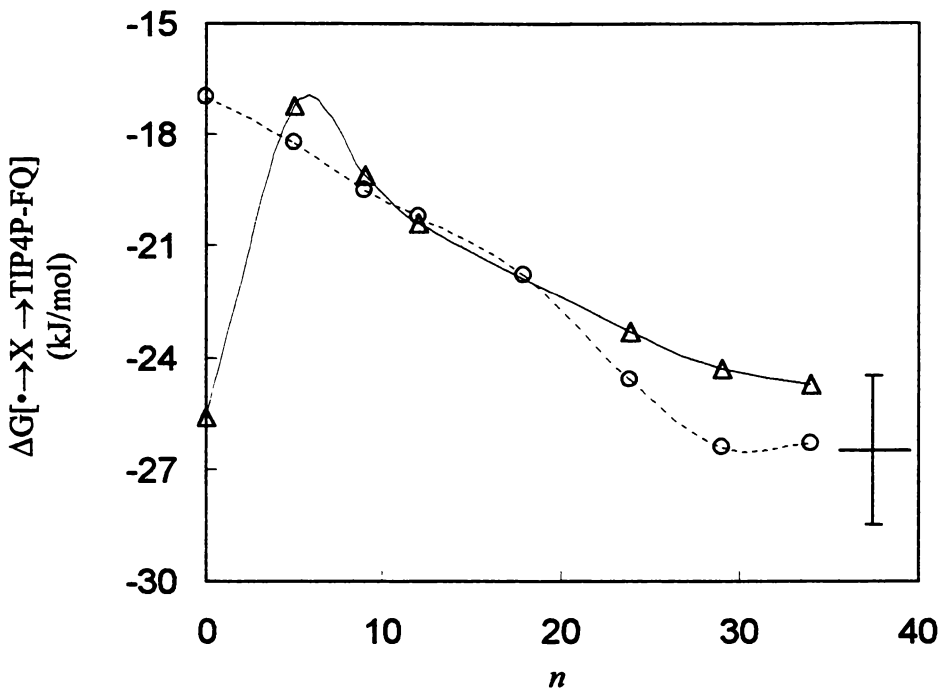

Figure 5.5 : Free energy of hydration using the M4 and TIP4P classical solute models and transforming the closest *n* solvent molecule interactions.
(O): ΔG[• → M4→FQ] ; (Δ): ΔG[• → TIP4P→FQ] ; The crosshairs denote the actual limiting free energy of hydration value and estimated uncertainty for the TIP4P-FQ model.

Table 5.4: Test of the ABC-FEP method on $M1 \rightarrow FQ$ and $TIP4P \rightarrow FQ$ as a function of n with $N_{conf} = 50$.

N	$\Delta G[M4 \rightarrow FQ_n]^*$ (kJ/mol)	$\Delta G[\bullet \rightarrow M4 \rightarrow FQ_n]$ (kJ/mol)	$\Delta G[TIP4P \rightarrow FQ_n]^*$ (kJ/mol)	$\Delta G[\bullet \rightarrow TIP4P \rightarrow FQ_n]$ (kJ/mol)
0	0.0	-17.0 ± 1.6	0.0	-25.6 ± 1.9
5	-1.2 ± 0.6	-18.2 ± 1.7	8.4 ± 0.4	-17.2 ± 1.9
9	-2.5 ± 0.3	-19.5 ± 1.6	6.5 ± 0.5	-19.1 ± 2.0
12	-3.2 ± 0.4	-20.2 ± 1.6	5.2 ± 0.7	-20.4 ± 2.0
18	-4.8 ± 0.6	-21.8 ± 1.7	-----	------
24	-7.6 ± 2.2	-24.6 ± 2.7	2.3 ± 0.7	-23.3 ± 2.0
29	-9.4 ± 4.0	-26.4 ± 4.3	1.3 ± 1.2	-24.3 ± 2.2
34	-9.3 ± 3.8	-26.3 ± 4.1	0.9 ± 1.4	-24.7 ± 2.4

The estimated errors are
$\pm Max\{-kT\log[\langle\exp(-\Delta U/kT)\rangle \pm \delta\langle\exp(-\Delta U/kT)\rangle] - [-kT\log\langle\exp(-\Delta U/kT)\rangle]\}$
where $\delta\langle\exp(-\Delta U/kT)\rangle$ is the statistical error of $\exp(-\Delta U/kT)$. This estimate does not include possibly large systematic errors when the peak in the exponentially weighted histogram is not well sampled.

kJ/mol. This is far from the desired value because M4 is not a *good* model for the interactions with the outer solvation shells. The transformation of the nearest five solvent waters from M4-FQ interactions to FQ-FQ interactions ($n = 0 \rightarrow 5$) essentially did not change $\Delta_s G$ as expected from the fact that M4 is adjusted to give minimum energy difference from TIP4P-FQ for the first solvation shell. However, as n continues to increase, $\Delta_s G$ continues to decrease until it finally converges to within 0.2 kJ/mol of the TIP4P-FQ model's calculated value at $n = 29$. The TIP4P model, on the other hand, yields -25.6 ± 1.9 kJ/mol for $n = 0$. When the interactions with the first five solvent waters are transformed from TIP4P to TIP4P-FQ interactions ($n = 0 \rightarrow 5$), $\Delta_s G$ shows a large increase (≈ 8.4 kJ/mol) signaling that TIP4P does not represent TIP4P-FQ water well in the first solvation shell. As the value of n continues to increase the value of $\Delta_s G$ decreases until it converges to within 2.0 kJ/mol of the TIP4P-FQ value at $n \geq 29$. For these two classical solute models (TIP4P and M4), the present method with $N_{conf} = 50$ gives the correct free energy of hydration of water as calculated directly from the H^b Hamiltonian (± 2.0 kJ/mol).

5.6 Full Quantum Example : Pure water at Supercritical Conditions

We have also applied the ABC-FEP method to calculating the free energy of hydration of water at 793.15K and 600 kg/m^3 using a full quantum treatment. This state point, while definitely in the supercritical region, is representative of some of the most extreme conditions possible in a hydrothermal vent. In addition, direct experimentation

on aqueous organics or even inorganic salts at these conditions are extremely difficult. Likewise, an accurate parameterization of classical potential models based on experimental data is often not possible. Even extrapolations using molecular dynamics and Monte Carlo simulations (parameterized at less extreme conditions) are likely to contain significant errors due to the non-transferability of the potential models to other non-parameterized state points. *Ab-initio* techniques, however, can offer much more accurate high temperature predictions because they already include the correct temperature dependencies in their formulations. This being said, the prediction of the free energy of hydration for water at 793 K and 600 kg/m^3 is very stringent test of the ABC-FEP method. Not only is water one of the most difficult molecules to model, very accurate experimental data in the supercritical region exists and can be used as a direct check of the ABC-FEP predictions.

In our earlier preliminary tests, we have shown that energy calculations with the cluster approximations gave reliable results when the perturbing cluster size was large enough (Wood *et al.* 1999); however, *ab initio* calculations of such large systems can be very expensive. In order to further reduce the computational costs, we have approximated the interaction energy of the solute with n solvent molecules as a sum of pair-wise and multi-body interaction energies,

$$U_{XS} = U_{XS,n}[pair] + U_{XS,m}[multi], \tag{5.13}$$

where the pair-wise interaction energy in a larger cluster n is a sum of all the solute (X) – solvent (S_i) pair interaction energies up to n, $U_{XS,n}[pair] = \sum_i^n u_{X,S_i}$, and the multi-body interaction energy is calculated as the difference between the pair-wise and total

interaction energies of a smaller cluster with m solvent molecules, $U_{XS,m}[multi] = U_{XS,m} - U_{XS,m}[pair]$. The full cluster interaction energy U_{XS} (or the pair interaction energy u_{X,S_i}) can be obtained from a total energy calculation of the solute X plus solvent S (or S_i) by subtracting the energy of isolated X and the total energy of the system in exactly the same solvent configuration without the solute particle (or the energy of isolated S_i). Thus the cluster method with $m < n$ can greatly reduce the amount of computational time since the pair-wise interaction energy calculations scale as N, where N is the number of dimer pairs, instead of $N^{3.5}$, where N is the total number of basis set functions.

The convergence of the free energy calculations as n and m increase depends on the *differences* in the quantum and classical energies ($\Delta U = U_{XS}^{Q} - U_{XS}^{C}$), *not* on the magnitude of the individual term, U_{XS}^{Q} or U_{XS}^{C}. We find that the energy differences converge much faster than the individual absolute classical or quantum intermolecular interaction energy terms, especially when the classical solute model is *good*. We also note that the cluster method can create unwanted surface potentials if the method of choosing the closest n molecules are not done correctly, especially if the system posses a net charge (Ashbaugh and Wood, 1997; Hummer *et al.*, 1997). However the influence of the surface potential to the interaction energies of a neutral solute is negligible, so we have used oxygen as a molecular center for convenience when choosing the closest n water molecules.

Also as one might expect the accuracy of our results will only be as good as the quality of the quantum mechanical method we employ. Thus we have decided to use

three different types of quantum mechanical calculations in order to give us a better understanding of the accuracy of the ABC-FEP method. The *ab initio* methods tested are i) density functional theory (DFT) using a B3LYP hybrid functional with the aug-cc-pVDZ basis set, ii) local Møller-Plesset perturbation theory at second order level (LMP2) (Sæbø and Pulay, 1986) with aug-cc-pVTZ, and iii) the SAPT five sites model (SAPT-5S) (Groenenboom *et al.*, 2000). SAPT-5S is an analytic function model that reproduces the water dimer interaction energy as a function of the relative positions of the two molecules keeping the monomer geometry fixed. All the DFT calculations are done with Gaussian 94, LMP2 was implemented with Jaguar 3.5 (1998) and SAPT-5S energies were calculated with the software of Mas and Szalewicz. All the quantum energies were corrected for basis set superposition error (BSSE) using the counterpoise method (Boys and Bernardi, 1970) (except for SAPT-5S which does not introduce BSSE). To reduce the computational cost of BSSE calculations, mean BSSE values are calculated by averaging over 5 configurations and then subtracted from each uncorrected interaction energy. This is valid since the standard deviation of BSSE is much smaller than that of the interaction energy (*e.g.*, the standard deviation of BSSE for DFT pair energies $U_{XS,60}^{DFT}[pair]$ in 5 configurations is 0.8 kJ/mol, while the standard deviation of $U_{XS,60}^{DFT}[pair]$ is 14.1 kJ/mol). We have estimated the error in $\Delta_s G$ introduced by using this mean BSSE in place of the individual BSSE for each configuration, and found it to be small (less than ± 2.0 kJ/mol for LMP2 and ± 1.0 kJ/mol for DFT).

Due to its well-documented performance, we have chosen the TIP4P-FQ as the classical solvent water. The first step was to calculate classical free energy of growing a

point mass to a full solute, $\Delta G[\bullet \rightarrow C] = 1.7 \pm 1.5$ kJ/mol. The second step is to determine $\Delta G[C \rightarrow Q]$ by examining $\Delta U_{XS} = U_{XS}^{Q} - U_{XS}^{C}$ via the cluster approximation. Table 5.5 include the mean pair-wise interaction energies for different methods with n = 5 to 42, and the mean multi – body interaction energies for TIP4P-FQ and DFT with m = 5 and 10. The results show that the multi – body contribution is negligible for both n = 5 and 10. The DFT energy is less negative than the TIP4P-FQ model by about 12 kJ/mol, while the LMP2 and SAPT-5S are with in 3 kJ/mol of the TIP4P-FQ value.

Using our previous criteria, all the $\langle \Delta U/kT \rangle$ and $\sigma[\Delta U/kT]$ values show that the TIP4P-FQ model gives *good* sampling for all the *ab initio* methods with N_{conf} = 200 (N_{conf} = 50 for LMP2). Figure 5.6 shows that the exponentially weighted distribution of $\Delta U/kT$ for a) DFT/DFT and c) SAPT-5S/DFT have completed distributions, but the *exp[-ΔU /kT]* distribution for b) LMP2/DFT is somewhat questionable due to much smaller number of samples (N_{conf} = 50). However, the estimated errors due to lack of samples are small for all the quantum methods studied here. The solvation free energies with DFT/DFT, LMP2/DFT, and SAPT-5S/DFT are plotted as a function of number of solvent molecules in the cluster (Figure 5.7). The empirical hydration free energies taken from Hill's equation of state for water (1990) at this condition yields, $\Delta_s G^{EOS}$ = -2.9 kJ/mol. Figure 5.7 and Table 5.6 show that the hydration free energy predicted with LMP2/DFT and SAPT-5S/DFT are converged to the empirical value ($\Delta_s G^{EOS}$) within the statistical error. The LMP2/DFT results give a slightly less negative value than the SAPT-5S/DFT results. The ABC-FEP method also successfully yields the correct $\Delta_s G$ to within 2 kJ/mol when LMP2/DFT or SAPT-5S/DFT is used although the classical free energy of

Table 5.5: Water interaction energies U_{XS} at 973.15 K and 600 kg/m^3 (kJ/mol)*

N	FQ		DFT		LMP2	SAPT-5S
	$<U_{XS,n}[pair]>$	$<U_{XS,n}[multi]>$	$<U_{XS,n}[pair]>$	$<U_{XS,n}[multi]>$	$<U_{XS,n}[pair]>$	$<U_{XS,n}[pair]>$
5	-16.7 ± 0.8	-0.2 ± 0.1	-10.6 ± 0.9	0.1 ± 0.1	-18.3 ± 1.8	-19.7 ± 0.9
10	-20.5 ± 0.9	0.2 ± 0.3	-11.6 ± 0.9	-0.1 ± 0.4	-22.2 ± 1.9	-24.0 ± 0.9
20	-23.0 ± 0.9		-12.2 ± 0.9		-24.8 ± 1.9	-26.0 ± 0.9
30	-23.7 ± 0.9		-12.3 ± 0.9		-24.8 ± 2.0	-26.2 ± 0.9
40	-24.1 ± 0.9		-12.4 ± 0.9		-25.5 ± 1.9	-26.3 ± 0.9
42	-24.2 ± 0.9		-12.5 ± 0.9		-25.5 ± 1.9	-26.4 ± 0.9

* all the energies are averaged over 200 independent configurations except $<U_{XS,n}[multi]>$ with $n = 10$ ($N_{conf} = 35$).

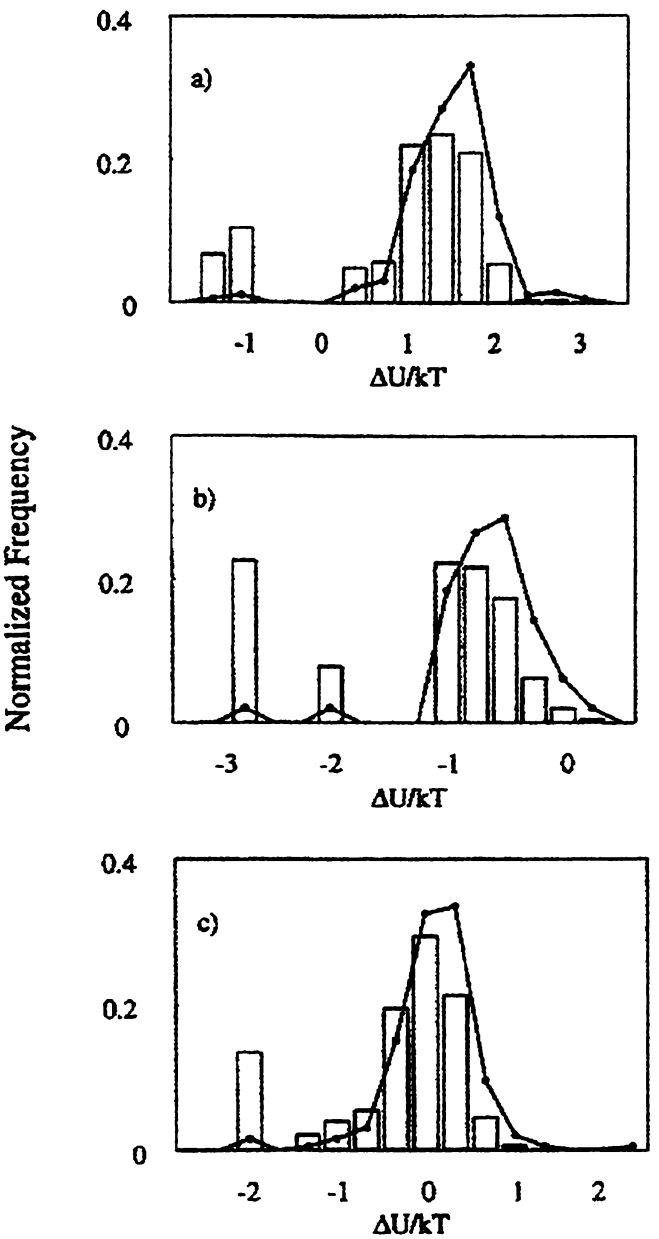

Figure 5.6 : Normalized ΔU_{XS} energy distributions at 973.15 K and 600 kg/m^3.
a) *DFT/DFT* (N_{conf} = 200), b) *LMP2/DFT* (N_{conf} = 50), c) *SAPT-5S/DFT* (N_{conf} = 200)

Solid line: $\Delta U_{XS}/kT$ distribution (N_{bin}/N_{conf})

Histogram: $\sum_{i=1}^{N_{bin}} \exp[-\Delta U_i/kT] \Big/ \sum_{j=1}^{N_{conf}} \exp[-\Delta U_j/kT]$

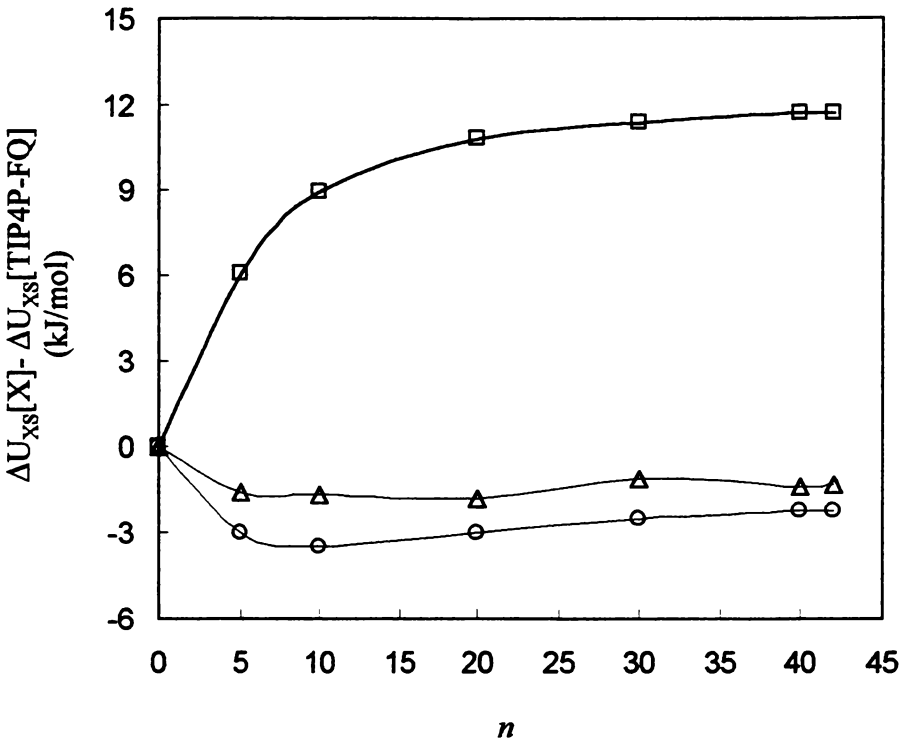

Figure 5.7 : Differences in solute-solvent interaction energy as functions of cluster size, *n*, at 973.15K and 600 kg/m^3. (□) *DFT/DFT*; (Δ) *LMP2/DFT*; (O) *SAPT-5S/DFT*. The interaction energies of the TIP4P-FQ were taken as the zero valued reference point.

Table 5.6: ΔU and ΔG results for pure water at 973.15K and 600 kg/m^3.

quantum method[a,b] *pair/multi*	n	m	N_{conf}	$\langle \Delta U/kT \rangle$	$\sigma(\Delta U/kT)$	$\Delta G[C \to Q]^*$ (kJ/mol)	$\Delta_s G^c$ (kJ/mol)
DFT/DFT	20	0	200	1.3 ± 0.1	0.5	8.9 ± 0.8	10.6 ± 1.7
	30	0	200	1.4 ± 0.1	0.6	9.5 ± 0.8	11.2 ± 1.7
	40	0	200	1.4± 0.1	0.5	9.8 ± 0.8	11.5 ± 1.7
	42	0	200	1.4 ± 0.1	0.6	9.8 ± 0.8	11.5 ± 1.7
	42	5	200	1.5 ± 0.1	0.6	10.0 ± 0.9	11.7 ± 1.7
LMP2/DFT	42	0	50	0.1 ± 0.1	0.6	-1.8 ± 2.0	-0.1 ± 2.5
	42	5	50	0.1 ± 0.1	0.7	-2.3 ± 2.8	-0.6 ± 3.2
SAPT-5s/DFT	30	0	200	-0.3 ± 0.1	0.5	-4.1 ± 0.7	-2.4 ± 1.7
	40	0	200	-0.3 ± 0.1	0.6	-3.8 ± 0.6	-2.1 ± 1.6
	42	0	200	-0.3 ± 0.1	0.5	-3.8 ± 0.6	-2.1 ± 1.6
	42	5	200	-0.2 ± 0.1	0.6	-3.6 ± 0.7	-1.9 ± 1.7

[a] $\Delta_s G^{EOS} = -2.9$ kJ/mol (Hill, 1990)

[b] $\Delta G[\bullet \to C] = 1.7 \pm 1.5$ kJ/mol

[c] $\Delta_s G = \Delta G[\bullet \to C] + \Delta G[C \to Q]$

*The estimated errors are

$$\pm Max\left\{-kT\log[\langle \exp[-\Delta U/kT]\rangle \pm \delta\langle \exp[-\Delta U/kT]\rangle] - [-kT\log\langle \exp[-\Delta U/kT]\rangle]\right\}$$

where $\delta\langle \exp[-\Delta U/kT]\rangle$ is the statistical error of $\exp[-\Delta U/kT]$. This estimate does not include possibly large systematic errors when the peak in the exponentially weighted histogram is not well sampled.

hydration ($\Delta G[\bullet \rightarrow C]$) is 4.6 kJ/mol off from $\Delta_s G^{EOS}$. However, with DFT/DFT, $\Delta_s G$ converges to the wrong value (11.7 ± 1.7 kJ/mol) and is presumed to be the result of an underestimation of the intermolecular dispersion forces.

An additional feature of the ABC-FEP method is that it is easy to find where the classical model is not reliable. In Figure 5.6, there are three $\Delta U/kT$ points which are far from the completed distributions. These three configurations were investigated in order to determine what was causing the large interaction energy differences between TIP4P-FQ and *ab initio* methods. We found that all three of configurations included a hydrogen bonding network around the solute water, and this network is pushing an oxygen atom of one solvent molecule to the solute oxygen atom so that the O – O distance is about 0.2 to 0.3.Å shorter than in typical configurations. This large energy difference is probably due to the TIP4P-FQ model's overestimation van der Waals energy on the repulsive wall (short distances). It was not necessary, however, to refine the TIP4P-FQ solute model to obtain accurate $\Delta_s G$ since the contribution of these three samples out of 200 configurations was small.

5.7 Conclusions

We have presented an accurate and computationally efficient method of calculating the free energy of solvation by perturbing the solute-solvent interactions from classical to quantum energies. For a test case of water at ambient conditions, different classical solute models gave the same free energy of hydration of water to within 2 kJ/mol. Using only 50 independent configurations generated by a classical simulation, a

bad classical model was easily recognized and transformed to a *good* model by two adjustment cycles. In our test case with the cluster model approximation, the rate of convergence as a function of n varies with the accuracy of the model. A *good* model gave the correct answer (± 2.0 kJ/mol) only with $n = 5$ ($\Delta G[\bullet \rightarrow (TIP4P - FQ) \rightarrow DFT_5]$). With defective models that are not accurate in the first shell (TIP4P) or the outer shells (M4), accurate results were achieved with $n \geq 29$. As a more stringent test, the free energy of hydration at 973.15 K and 600 kg/m^3 was accurately calculated using a full quantum mechanical single perturbation. In addition the ABC-FEP method has the potential of being applicable to a wide variety of other free energy calculations: chemical potentials in mixtures, solvent effects on the activation barriers, the free energies of a pure fluids, and free energies of "solute" transformation on a surface or in a solid.

References

Allen, M. P. and Tildesley, D. J. 1989. *Computer Simulation of Liquids*, Clarendon Press, Oxford.

Ashbaugh, H.S. and Wood, R. H. 1997. *J. Chem. Phys.*, **106**, 8135.

Ben-Naim, A. and Marcus, Y. 1984. *J. Chem. Phys.*, **81**, 2016.

Berendsen; H.J.C.; Grigera, J. R. and Straatsma, T. P. 1987. *J. Phys. Chem.*, **91**, 6269.

Boys, S. F. and Bernardi, F. 1970. *Mol. Phys.*, **19**, 553.

Chialvo, A.A. and Cummings, P. T. 1996. *J. Chem. Phys.*, **105**, 8274.

Curtis, L. A.; Frurip, D. J. and Blander, M. J. 1979. *Chem. Phys.*, **71**, 2703.

Dyke, T. R. and Muenter, J. S. 1973. *J. Chem. Phys.*, **59**, 1325.

Field, M. J.; Bash, P. A. and Karplus, M. 1990. *J. Comp. Chem.*, **11**, 700.

Frenkel, D. and Smit, B. 1996. *Understanding molecular simulation: From algorithms to applications.* Academic Press, New York.

Forester, T. R. and Smith, W. 1995. DL_POLY_2.0 Daresbury Laboratory, Daresbury: Warrington, England.

Gao, J. 1992. *J. Phys. Chem.* **96**, 537.

GAUSSIAN 94 1995. Gaussian, Inc., Pittsburgh, PA.

Groenenboom, G. C.; Mas E. M.; Bukowski, R.; Szalewicz, K.; Wormer, P.E. S. and van der Avoird. 2000. *Phys. Rev. Lett.*, **84**, 4072.

Hill, P. G. 1990. *J. Phys. Chem. Ref. Data*, **19**, 1233.

Hummer, G.; Pratt, L. R.; Garcia, A. E.; Berne, B. J. and Rick,S.W. 1997. *J. Phys. Chem. B*, **101**, 3017.

Hummer, G. and Szabo, A. 1996. *J. Chem. Phys.*, **105**, 2004.

Jaguar 3.5 1998. *Schrödinger, Inc.*, Portland, OR.

Kim, K. and Jordan, K.D. 1994. *J. Phys. Chem.*, **98**, 10089.

Laasonen, K. and Klein, M. L. 1996. *Mol. Phys.*, **88**, 135.

Lin, C. and Wood, R. H. 1996. *J. Phys. Chem.*, **100**, 16399.

Mas, E. M. and Szalewicz, K. 1997. *J. Chem. Phys.*, **107**, 4207.

McQuarrie, D.A. 1976. *Statistical Mechanics*, Harpers Collins Publishers, New York.

Mezei, M. 1991. *J. Phys. Chem.*, **95**, 7042.

Rick, S.W. and Berne, B. J., 1996. *J. Am. Chem. Soc.*, **118**, 672.

Rick, S.W.; Stuart, S. J. and Berne, B. J. 1994. *J. Chem. Phys.*, **101**, 6141.

Sæbø, S. and Pulay, P. 1986. *Theor. Chim. Acta*, **69**, 357.

Sakane, S.; Ashbaugh, H. and Wood, R. H. 1998. *J. Phys. Chem. B*, **102**, 5673.

Sanderson, R. T. 1951. *Science*, **114**, 670.

Sprik, M.; Hutter, J., and Parrinello, M. 1996. *J. Chem. Phys.*, **105**, 1142.

Tomasi, J. and Persico, M. 1994. *Chem. Rev.*, **94**, 2027.

Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W.A. and Honig, B. 1994. *J. Am. Chem. Soc.*, **116**, 11875.

Wood, R. H.; Mühlbauer, W.C.F. and Thompson, P.T. 1991. *J. Phys. Chem.*, **95**, 6670.

Watanabe, K. and Klein, M. L. 1989. *J. Chem. Phys.*, **131**, 157.

PART 6

ELUCIDATING THE ROLE HYDROTHERMAL SOLUTIONS PLAY IN INORGANIC/ ORGANIC EQUILIBRIUM

6.1 Insights into Hydrothermal Organic Chemistry

At this time it would be constructive to review some of our results and directly relate them to hydrothermal chemistry. The EOS presented in part 2 represents one of the most accurate and wide ranging descriptions of dilute organic aqueous equilibrium, to date. The only other comparable work was performed by Amend and Helgeson (1997) and employed the revised HKF equation of state (Helgeson *et al.*, 1981; Shock and Helgeson, 1988,1990). However because their findings were published before the recent high temperature organic data became available, and because the HFK equation does not give an accurate description in the critical region (O'Connell *et al.*, 1996), the maximal temperature of their predictions was limited to 520 K. The parameterization of the equation presented in part 2 (Yezdimer *et al.*, 2000; Sedlbauer *et al.*, 2000) however, can be expected to yield accurate chemical potentials up to 625 K, and does appear to express the correct limiting critical and infinite temperature trends.

As was demonstrated in part 2 accurate knowledge of the free energy of hydration is of fundamental importance when attempting to model the total free energy of a solute in solution. The free energy of formation for a single solute in solution can be given by,

$$\Delta G_f = \Delta G_2^{id} + \Delta G_2^{hyd} + \textit{Standard State Terms} \qquad (6.1)$$

The quantity ΔG_2^{id} is simply the gas-phase free energy and may typically be determined through *ab-initio* methods, tabulated spectroscopy data or, as it was used in part 2, by experimental heat capacity data. The contribution from all the solute-solvent interactions is contained in the ΔG_2^{hyd} term and is considerably more challenging to model accurately. However, the good agreement between the molecular dynamic results presented in part 4

and the Henry's constants predicted by Yezdimer *et al.* (2000), particularity the reversal of the solubility trends with increasing temperature, supports not only the EOS's ΔG_2^{hyd} predictions, but also the EOS's description of the free energy of formation. Similar qualitative solubility trends for several hydrocarbons were also independently obtained by Plyasunov and Shock (2000), through a valuation of experimental hydrocarbon/water VLE and LLE data.

The conformation of the EOS's predicted trends for a homogeneous aqueous organic series has important implications for the hydrothermal origin of life theory. This can be seen predominately in the predicted values for the free energy of formation, ΔG_f (Figure 6.1). While the equations of Yezdimer *et al.* (2000) predict that larger organic molecules (*i.e.*, octane) are less stable than smaller molecules (*i.e.*, methane) at room temperature, the situation appears to be reversed at high temperatures. This behavior in the ΔG_f curves was first reported by Amend and Helegson (1997), however, for reasons discussed earlier they could only speculate as to the accuracy of their predictions at those temperatures and implied the behavior could be unphysical. Although it now appears that the addition of a CH_2 group does generally serve to decrease the free energy of a given organic molecule at elevated temperatures (Yezdimer *et al.*, 2001a,b; Plyasunov and Shock, 2000), instead of increasing it (which is the case at room temperature).

These findings generally tend to support the hydrothermal origin of life theory, or at least a hydrothermal origin for the creation of a localized source of organic material.

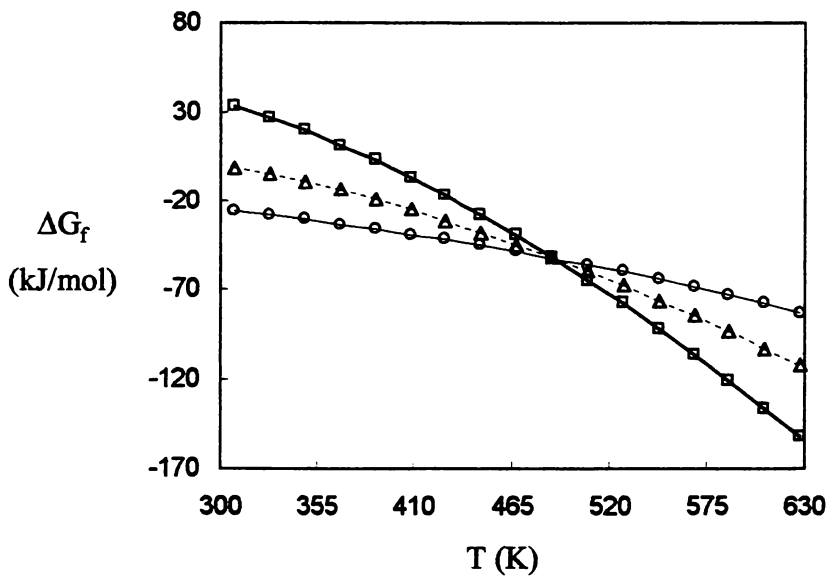

Figure 6.1: Free energies of formation for a series of aqueous alkanes. The symbols denote methane(aq), O ; butane(aq), Δ ; and octane(aq), □.

Simply put, the existence of larger, more complex organic molecules appears to be energetically *preferred* over the existence of smaller, less complex species. This conclusion is quite different from simply predicting that the free energy of formation for a particular organic molecule will decrease with increasing temperature, but rather it states that larger organic solutes in solution will exist at a thermodynamically lower free energy state than smaller organic solutes. Shock and Schulte (1998) have reported a similar preference for the formation of large organic molecules over that of small organic molecules. The results from their hydrothermal equilibrium models suggest that the distributions of organic molecules formed from the reaction $a\ CO_2 + b\ H_2 \rightarrow organic + d\ H_2O$ are dominated by larger organic molecules at high temperatures. Because of the ubiquitous nature of the CH_2 group the reversal in free energy trends at high temperature are by no means simply limited to hydrocarbons. Alcohols, amines, amides, diols, diamides, diamines, and even carboxylic acids all are predicted to show similar trends with increasing molecular size (Figure 6.2). Although, it would be extremely interesting to see if a cyclic CH_2 functional group also showed the same transition at elevated temperatures in its free energy contribution (from positive to negative) as is true for the acyclic CH_2 group.

Although the reversal in the free energy of formation trends at high temperature may seem novel at first, the hindsight explanation for the reversal is quite trivial. By a simple magnitude analysis of each term in Eqn. (6.1) it is clear that the dominating contribution to the free energy of formation is from the ideal gas component. This is not too surprising because one might expect the free energy of formation to be a much

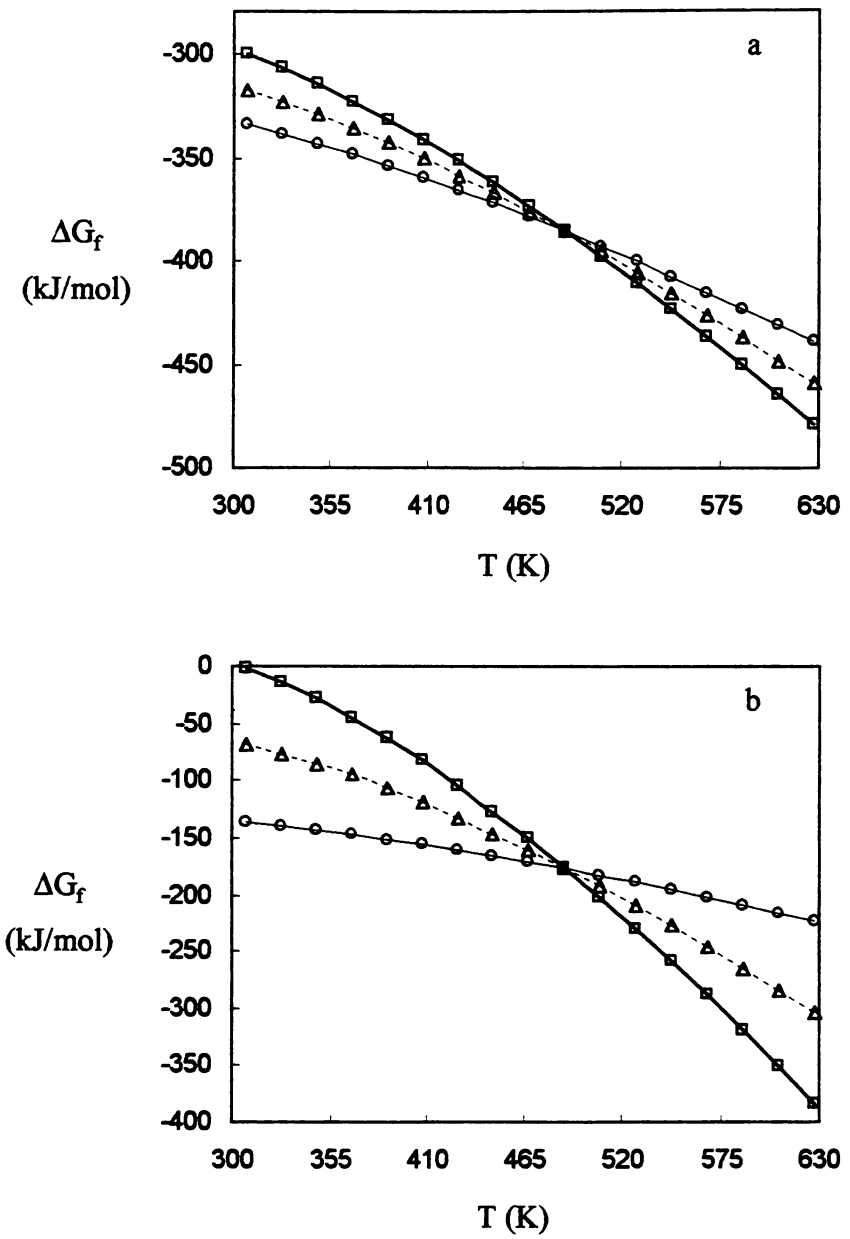

Figure 6.2: Free energies of formation for several polar and branched compounds. a) Series of 1,n-diols with *n* number of carbon atoms, *n*=3(aq), O ; *n*=5(aq), Δ ; and *n*=7(aq), □. b) Series of alcohols with periodic methyl branching, HO-(-$CH(CH_3)CH_2CH_2)_n$-CH_3, *n*=1(aq), O ; *n*=3(aq), Δ ; and *n*=5(aq), □.

stronger function of the intramolecular structure and energy than the much weaker intermolecular interactions. Larger molecules translate into larger gas-phase entropies and result in an overall preference of the larger, more complex molecules at high temperatures.

Despite the dominance of the ideal gas free energy term in Eqn. (6.1), it interesting that the free energy of hydration term also shows a similar energetic preference for the existence of larger organic molecules at high temperatures (see part 4). Clearly water appears to be a more comprehensive solvent at temperatures over 473 K, allowing even some of the most hydrophobic species (*i.e,* saturated hydrocarbons) to exist in the aqueous state. Taken in conjunction with the molecular size dependence of the ideal gas energy term, it does appear that hydrothermal solutions represent a much more ideal environment for the synthesis and chemical evolution of large complex organic molecules than water at ambient conditions. This is not to say, however, that inorganic to organic equilibria fall into the spontaneous region under hydrothermal conditions (only a further investigation into the thermodynamics of the inorganic source material and resulting chemical pathways could revel the actual equilibrium constants). Although it does indicate that at least the product's side of the chemical reaction will favor the formation of large organic molecules.

From a modeling perspective, the solvent contribution to the free energy of a solvated molecule is much more difficult to describe accurately. The magnitude of the free energy of hydration for most neutral organic molecules is roughly between 3 – 7 thermal units and can result in a several order of magnitude in the calculated equilibrium constants, if not treated properly. For organic electrolytes, the need for an accurate

determination of the free energy of hydration is even more necessary due to a large amount of solvent rearrangement resulting from their ionic nature. One of the most fundamental problems in modeling the solvent effects is dealing with the strong temperature and density dependence of the polar character of water. From the results of classical modeling techniques outlined in part 3, it is clear that no single model can properly describe the microstructure, thermodynamics, and phase behavior of pure bulk water simultaneously. Problems with selecting the most "realistic" combing rules when modeling dilute mixtures also serve to increase the modeling difficulties.

In an effort to tackle these obstacles, the ABC-FEP method was specifically developed and is all ready beginning to prove extremely useful. For instance, Wood and co-workers (2001) have recently been able to use the ABC-FEP method to calculate out the combined free energy of hydration for the Na^+ and Cl^- ions and their results agreed very well with the accepted experimental value, within 1 part in 300. Applying the ABC-FEP method to several key organic compounds as a function of temperature (up to T=1000 K) should provide valuable and quantum mechanical level insights as to the interplay between solvent's dielectric/structure/polarizable nature and favorable changes in the system's free energy for large dilute organic molecules.

It is also rather interesting to note the continuous nature of the organic solute free energy curves in the vicinity of the critical point. Most other thermodynamic properties (*i.e.*, partial molar volumes, heat capacities, enthalpies, and entropies) all show a divergent behavior at the critical point. The reason for this behavior on the microscopic level has been suggested by Chialvo and Cummings (1994), who were able to show that the free energy of solvation should be independent of any long range, compressibility

driven behavior. The same however cannot be said of the other thermodynamic properties, which are dominated by long-range correlations near the critical point. Similarly one might also expect this behavior to limit the applicable range of the functional group addivity assumption for the divergent properties. Unfortunately, given the lack of experimental and theoretical data in the near critical region, it is very difficult to make any kind of definite determination as to the accuracy of the functional group approach in the critical region.

6.2 Directions for Future Work

Despite the advances outline thus far, several major stumbling blocks still remain in our treatment of hydrothermal conditions. On the molecular level, one of the most challenging problems lies in properly modeling the thermodynamic and structural properties of liquid water (in addition to their temperature and density dependencies). While classical polarizable, and even non polarizable, models can provided a good description of several basic bulk thermodynamic properties (*i.e.* intermolecular interaction energy, density, heat of vaporization, dielectric constant, viscosity), the accurate prediction of other properties (*i.e.* free energies, dipole moments, phase coexistence lines) still remains elusive (Yezdimer and Cummings, 1999; Chialvo *et al.*, 2000). The modeling of solute-solvent interactions is also a major limiting factor in the accuracy of molecular level based approaches (*i.e.* molecular dynamics or Monte Carlo simulations). Significant errors in the predicted thermodynamics, as well as the structural properties, are often introduced either from an empirically derived generalized force

fields or by the use of approximate combining rules, as demonstrated in part 4 (Yezdimer *et al.*, 2001a,b).

Methods like the ABC-FEP (Wood *et al.*, 1999; Sakane *et al.*, 2000a,b) are beginning to prove themselves invaluable in making predictions of aqueous chemical potentials, particularly in the supercritical region where experimentation is very difficult and the lower densities in the supercritical region also make them computationally ideal for a quantum mechanical treatment. The eventual hope is that the ABC-FEP method (or other similar techniques) will be able to provide highly accurate (2-4 kJ/mol) chemical potential data that could then be combined with direct experimental data obtained at lower temperatures, in order to give a clear understanding of the systems thermodynamics over the entire temperature range.

Another major difficulty to provide the microscopic description of a hydrothermal system is the rather large amount of computational resources required to execute calculations at the molecular dynamics level. Even when today's state-of-the-art computers are employed, a full treatment using the ABC-FEP method would take at least one month of CPU time to investigate a single organic compound at only a single state condition. Although, this is a vast improvement from previous *ab-initio* techniques for treating bulk systems (which would take on the order of centuries), the calculations are not fast enough to allow for a compound-by-compound description of a complex hydrothermal system. In addition, the slowness of the approach is not entirely due to the cost of the *ab-initio* calculations. Current classical free energy calculations typically required over 2 million computational steps in order to obtain accurate free energy of solvation for a simple organic molecule at infinite dilution. Unfortunately the

computational speed of current techniques scale anywhere between N^2 and $N \log N$, where N is the number of interactions sites in the system, resulting in a major draw on computational resources for the explicit evaluation of the solvent-solvent intermolecular distances. This fact has effectively limited the size of most molecular dynamics or Monte Carlo studies systems to only 500 to 1000 small molecules and timescales under 1-2 nanoseconds. As was demonstrated in part 5, accurate quantum mechanical free energies of solvation could be obtained by sampling solvent configurations from a more computationally efficient method. If a similar technique could be developed to remove/reduce the need for performing rigorous solvent-solvent distance calculations during the course of a classical molecular dynamics simulation, while maintaining an acceptable treatment of the solvent's Hamiltonian, it would constitute a tremendous breakthrough in classical computational chemistry and would represent one of the first steps toward the explicit modeling of complex chemical reactions occurring in solution.

On the macroscopic level, empirical based equations of states offer a much wider range of insight into organic/inorganic equilibrium at elevated conditions. The equation of state presented in part 2 has shown itself to be quite accurate both in terms of its quantitative and qualitative predictions of basic thermodynamic properties (*i.e.*, Henry's constant, partial molar volume data, Gibbs free energies of formation). However, equations of state approaches are limited by the amount of available experimental data, which is often scarce and difficult to obtain (especially for aqueous organics at high temperatures). The application of functional group addivity comes into question when dealing with the very complex organic molecules that are commonplace in biological systems. Although work is currently being done on expanding the range of available

experimental data to include ethers, esters, ketones, and multi substituted aromatics (Sedlbauer and Majer, In progress), much more additional experimentation will be required to provide a fully comprehensive model of primordial organic/inorganic chemistry.

Ideally the best future approach would consist of a hybrid between both the macroscopic EOS approach and the microscopic MD/QM techniques. Clearly the advantages of such a dual approach can be seen in the molecular dynamics predictions of ΔG^{hyd} for butane and octane (Yezdimer *et al.*, 2001a,b), where the molecular dynamics level approach was able to confirm the near critical solubility trends (where direct experimentation was unattainable). Moreover, a method where it would be possible to use either an empirically derived potential surface or a quantum mechanical developed surface to directly determine the EOS's adjustable parameters would prove extremely beneficial. Some progress towards this goal has been accomplished, namely the rigorous formulation of the free energy of hydration presented in section 4.2. Unfortunately a much firmer understanding behind the empirical EOS parameter's physical meaning, and a well-defined manner in which they could be determined, are required before the EOS parameters could be determined purely from *ab-initio* techniques.

In terms of the most interesting chemical system for future study, a mixture of $H_2O + CO_2$ + dilute organics may prove to be the most fruitful. This type of system is much more representative of an actual hydrothermal environment. The key point is that the mixture's critical point is likely to lie significantly below the critical point of pure water. If the solubility trends predicted in part 2 and 4 are indeed a result of the sharp decreases in solvent density (along a given isobar) at the critical point, the reversal in

solubility trends could appear at a much lower temperature. Such a shift in the solubility trends would in theory, allow for an increased number of possible chemical reactions involving much larger aqueous organic compounds than in a single component (H_2O) mixture. In terms of developing alterative industrial solvents, the effects on the critical properties of other less reactive additives should also be examined.

References

Amend, J. P. and Helgeson, H. C. 1997. *Geochim.Cosmochim.Acta*, **61**, 11.

Chialvo, A.A. and Cummings, P.T. 1994. *AIChE J.*, **40**, 1558.

Chialvo, A.A.; Yezdimer, E.M.; Driesner, T.; Cummings, P.T. and Simonson, J.M. 2000. *Chem. Phys.*, **258**, 109.

Helgeson, H. C.; Kirkham, D. H. and Flowers, G. C. 1981. *Am. J. Sci.*, **281**, 1249.

O'Connell, J. ; Sharygin, A.V. and Wood, R. H. 1996. *Ind.Eng. Chem. Res.*, **35**, 2808.

Plyasunov, A.V. and Shock, E.L. 2000. *Geochim. Cosmochim. Acta.*, **64**, 2811.

Sakane,S.; Lui, W.; Yezdimer, E.M.; Barriocanal, J.A.; Doren, D.J.; and Wood, R.H. 2000a. "High level calculations of water properties using first-principles with classical simulations," *Steam, Water, and Hydrothermal Systems: Physics and Chemistry Meeting the Needs of Industry.* Eds. Tremaine *et al.* NRC Press, Ottawa.

Sakane, S.; Yezdimer, E.M.; Liu, W.; Barriocanal, J.A.; Doren, D.J. and Wood, R.H. 2000b. *J. Chem. Phys.*, **113**, 2583.

Sedlbauer, J.; Yezdimer, E.M.; Majer, V. and Wood, R.H. 2000. "Correlation and prediction of standard thermodynamic properties of aqueous organic solutes in a wide range of conditions," *Steam, Water, and Hydrothermal Systems: Physics and Chemistry Meeting the Needs of Industry.* Eds. Tremaine *et al.* NRC Press, Ottawa.

Sedlbauer, J. and Majer, V. 2001. In progress.

Shock, E. L. and Helgeson, H.C. 1988. *Geochim. Cosmochim. Acta*, **52**, 2009.

Shock, E. L. and Helgeson, H.C. 1990. *Geochim. Cosmochim. Acta*, **54**, 915.

Shock, E. L. and Schulte, M.D. 1998. *J. Geophys. Res.*, **103**, 28513.

Wood *et al.* 2001. Private Communication.

Wood, R.H.; Yezdimer, E.M.; Sakane, S.; Barriocanal, J.A.and Doren, D.J. 1999. *J. Chem. Phys.*, **110**, 1329.

Yezdimer, E.M.; Chialvo, A.A. and Cummings, P.T. 2001. *Fluid Phase Equil.*, In Press.

Yezdimer, E.M.; Chialvo, A.A. and Cummings, P.T. 2001. *J. Phys. Chem. B* , **105**, 841.

Yezdimer, E.M. and Cummings, P.T. 1999. *Mol. Phys.* **97**, 993.

Yezdimer, E.M.; Sedlbauer, J.; and Wood, R.H. 2000. *Chem. Geol.* **164**, 259.

APPENDIX

Appendix A : Equation of State Examples

Test calculations are summarized in this appendix for two test solutes, propanoic acid and methylammonium propanoate at T = 348.15 K and P = 0.1 MPa. The results of these calculations can be used as a reference when implementing the group contribution scheme into a computer program. Thermodynamic properties of water were calculated from Hill's equation of state (Hill, 1990) and all properties, for water and for aqueous solutes, were employed in basic SI units. Standard state terms, Eqns.(2.9), (2.11) - (2.13) at this temperature and pressure are: $V_{SS} = 1.3 \cdot 10^{-6} m^3 \cdot mol^{-1}$, $\Delta_{hyd} H_{SS} = -2280 J \cdot mol^{-1}$, $\Delta_{hyd} S_{SS} = -67.7 J \cdot K^{-1} \cdot mol^{-1}$, $\Delta_{hyd} C_{P,SS} = 1.0 J \cdot K^{-1} \cdot mol^{-1}$ and conversion factor in Eqn.(2.1) $\Lambda_S = 33.4 J \cdot K^{-1} \cdot mol^{-1}$.

Example 1 – propanoic acid.

2 C groups, 5 H groups and 1 COOH group. Charge of the acid molecule is $z = 0$. Thermodynamic properties at reference conditions are calculated from the values in Table A1 and the thermodynamic properties of hydration at elevated conditions are found from relevant equations in the Appendix, using parameters of the groups from Table 2.1. The partial and final results of calculating standard thermodynamic properties are given in Table A2.

Table A1: Parameters for functional group scheme for $C_{p,2}^{ig}[T]$ at P^0= 0.1 MPa[a] and functional group contributions to S_2^0 and $\Delta_f H_2^0$ at T_r=298.15 K and P_r= 0.1 MPa[b].

	$\Delta_a \cdot 10^{-1}$	$\Delta_b \cdot 10^{1}$	$\Delta_c \cdot 10^{4}$	$\Delta_d \cdot 10^{7}$	S_2^0	$\Delta_f H_2^0$
	$J/K/mol$	$J/K^2/mol$	$J/K^3/mol$	$J/K^4/mol$	$J/K/mol$	J/mol
C	-4.425	3.1004	-4.74	2.2835	-105.16	29177
H	2.125	1.0604	2.09	-1.0835	65.46	-26557
$CONH_2$	3.35	0.258	1.283	-0.9474	160.28	-256326
$COO^{(-)}$	2.41	0.427	0.804	-0.687	129.17	-432419
COOH	2.41	0.427	0.804	-0.687	155.18	-434569
NH_2	2.69	-0.412	1.64	-0.976	95.27	-25123
$NH_3^{(+)}$	2.69	-0.412	1.64	-0.976	51.48	-74436
OH	2.57	-0.691	1.77	-0.988	98.40	-213023
AMINO	5.1	0.015	2.444	-1.663		

[a] Reid *et al.* (1987). Parameters for C and H groups were calculated from the parameters for C, CH, CH_2 and CH_3 groups. Parameters for the charged groups COO- and NH_3^+ were assumed to be the same as for the uncharged groups COOH and NH_2. Parameters for the AMINO aminoacid group are calculated as a sum of parameters for COO- and NH_3^+ functional groups.

[b] Calculated from formation properties reviewed by Amend and Helgeson (1997a) and Shock and Helgeson (1990), values for NH_3^+ group calculated from experimental results on methylammonium ion from Wagman *et al.* (1982). Standard state terms at T_r=298.15 K and P_r= 0.1 MPa: $\Delta_{hyd}S_{SS} = -67.74\ J \cdot K^{-1} \cdot mol^{-1}$, $\Delta_{hyd}H_{SS} = -2288\ J \cdot mol^{-1}$. The entropies of constituent elements in their standard states at T_r and P_r: $S_C^0[298K,0.1MPa] = 5.740\ J \cdot K^{-1} \cdot mol^{-1}$, $S_{H_2}^0[298K,0.1MPa] = 130.684\ J \cdot K^{-1} \cdot mol^{-1}$, $S_{O_2}^0[298K,0.1MPa] = 205.138\ J \cdot K^{-1} \cdot mol^{-1}$, $S_{N_2}^0[298K,0.1MPa] = 191.61\ J \cdot K^{-1} \cdot mol^{-1}$.

Table A2: Test calculations of standard thermodynamic properties of propanoic acid and methylammonium propanoate at T = 348.15 K and P = 0.1 MPa.

	propanoic acid	methylammonium propanoate
$10^6 \cdot \sum_{i=1}^{N} n_i V_{2,i}^{0}$ [a]	70.7	90.9
$10^6 \cdot V_2^{0}$ [a]	72.0	93.6
$C_{P,2}^{ig}$ [b]	99.5	159.0
$\sum_{i=1}^{N} n_i C_{P,2,i}^{0}$ [b]	154.2	89.2
$C_{P,2}^{0}$ [b]	254.7	250.2
$\Delta_f H_2^0[T_r, P_r]$ [c]	-511300	-636400
$\Delta_{hyd} H_2^0[T_r, P_r]$ [c]	-63000	-131100
$\int_{T_r}^{T} C_{P,2}^{ig} dT$ [c]	4700	7600
$\sum_{i=1}^{N} n_i \cdot \Delta_{hyd} H_{2,i}^{0}$ [c]	-52900	-121300
$\Delta_{hyd} H_2^{0}$ [c]	-55200	-125800
$\Delta_f H_2^{0}$ [c]	-498800	-623600
$S_2^0[T_r, P_r]$ [b]	204.4	253.4
$\Delta_{hyd} S_2^0[T_r, P_r]$ [b]	-126.8	419.3
$\int_{T_r}^{T} C_{P,2}^{ig} / T dT$ [b]	14.6	23.4
$\sum_{i=1}^{N} n_i \cdot \Delta_{hyd} S_{2,i}^{0}$ [b]	-68.2	504.3
$\Delta_{hyd} S_2^{0}$ [b]	-102.5	435.6
S_2^{0} [b]	243.3	293.1
$\sum_{el} S_{el}^0[T_r, P_r] - z/2 \cdot S_{H_2}^0[T_r, P_r]$ [b]	614.4	1042.7
$\Delta_f G_2^{0}$ [c]	-400300	-414700

[a] $m^3 \cdot mol^{-1}$

[b] $J \cdot K^{-1} \cdot mol^{-1}$

[c] $J \cdot mol^{-1}$

Example 2 – methylammonium propanoate.

3 C groups, 8 H groups, COO^- group and NH_3^+ group. Eqn.(2.1) has to be used for both constituting ions, with charges $z = 1$ for methylammonium ion and $z = -1$ for propanoate ion (thus, two standard state terms and two conversion factors are included in every calculated standard thermodynamic property of a salt with two ions). When calculating correction terms, Eqns.(A14) – (A16), the equations differ for nonelectrolyte groups C and H and for ionic groups COO^- and NH_3^+. The charge of a group also determines the value of δ_i, Eqn.(A7). The procedure is the same as in Example 1 and the results are again summarized in Table A2.

Appendix B : Review of Combining Rules

The pair potential due to dispersion interaction between to unlike Lennard-Jones spheres can be written as,

$$\varphi_{12}^{dis} = -\frac{3}{2}\left(\frac{E_{I1}E_{I2}}{E_{I1}+E_{I2}}\right)\frac{\alpha_1\alpha_2}{r_{12}^6} = -\frac{4\varepsilon_{12}\sigma_{12}^6}{r_{12}^6}, \tag{B.1}$$

where E_I, is the ionization potential, α is the species polarizablity [not to be confused with the Ewald parameter], r is the intermolecular distance between sphere centers, and ε and σ are the standard Lennard-Jones parameters. Thus the well depth of the Lennard-Jones curve can be expressed in the form,

$$\varepsilon_{12} = \frac{3}{8}\left(\frac{E_{I1}E_{I2}}{E_{I1}+E_{I2}}\right)\frac{\alpha_1\alpha_2}{\sigma_{12}^6} \tag{B.2}$$

The well depths of the like molecular Lennard-Jones interactions can described by,

$$\varepsilon_{11} = \frac{3}{16}\frac{E_{I1}\alpha_1^2}{\sigma_{11}^6} \tag{B.3}$$

$$\varepsilon_{22} = \frac{3}{16}\frac{E_{I2}\alpha_2^2}{\sigma_{22}^6} \tag{B.4}$$

The deviation from the Berthelot combining rule can then be calculated as,

$$\frac{\varepsilon_{12}}{\sqrt{\varepsilon_{11}\varepsilon_{22}}} = 2\left(\frac{\sqrt{E_{I1}E_{I2}}}{E_{I1}+E_{I2}}\right)\frac{\sigma_{11}^3\sigma_{22}^3}{\sigma_{12}^6} \tag{B.5}$$

Using the Lorentz combining rule, $\sigma_{12} = (\sigma_{11}+\sigma_{22})/2$, Eqn. (B.5) can be expressed as,

$$\frac{\varepsilon_{12}}{\sqrt{\varepsilon_{11}\varepsilon_{22}}} = \left(\frac{2\sqrt{E_{I1}E_{I2}}}{E_{I1}+E_{I2}}\right)\left(\frac{64\sigma_{11}^3\sigma_{22}^3}{(\sigma_{11}+\sigma_{22})^6}\right) \tag{B.6}$$

By a simple substitution of Eqn. (B.3) and Eqn. (B.4) back into Eqn. (B.6) one can then quickly arrive at what is referred to as the Kohler (1957) combining rules. Clearly in the case of $E_{I1} = E_{I2}$ and $\sigma_{11} = \sigma_{22}$, Eqn. (B.6) simply reduces to the Berthelot combining rule. Thus two additional factors can be added to the Berthelot combining rule. The first factor takes into account dissimilarities between ionization energies, and the second factor takes into account dissimilarities in molecular size. Also it is important to notice, as pointed out by Hildebrand and Scott (1962), that the product of both these additional factors will be always between 1 and 0 and hence the normal Berthelot combining rule will always overestimate the unlike pair interaction parameter. Other more complex combining rules have been developed (Smith, 1972; Kong, 1973; Lucas, 1986; Shukla, 1989), however the use of this simplified version of the Kohler combining rules is sufficient to illustrate the effect that differences in molecular size has on the free energy of hydration.

References

Amend, J. P. and Helgeson, H. C. 1997a. *Geochim.Cosmochim.Acta*, **61**, 11.

Hildebrand, J.H. and Scott, R.L. 1962. *Regular Solutions*, Prentice Hall In, Englewood Cliffs, New Jersey.

Hill, P.G. 1990. *J.Phys.Chem.Ref.Data*, **19**, 1233.

Kohler, F. 1957. *Monatshefte Chemie*, **88**, 857.

Kong, C.L. 1973. *J. Chem. Phys.*, **59**, 2464.

Lucas, K. 1986. *Applied Statistical Thermodynamics*, Springer-Verlag, Berlin.

Reid, R.C.; Prausnitz, J.M. and Poling, B.E. 1987. *The properties of gases and liquids*. McGraw-Hill Book Company, New York.

Shock E.L. and Helgeson H.C. 1990. *Geochim.Cosmochim.Acta*, **54**, 915.

Shukla, K.P. 1989. *Fluid Phase Equilibra*, **46**, 1.

Smith, F.T. 1972. *Phys. Rev.*, **A5**, 1708.

Wagman, D.D.; Evans, W.H.; Parker, V.B.; Schumm, R.H.; Halow, I.; Bailey, S.M.; Churney, K.L. and Nuttall, R.L. 1982. *J.Phys.Chem.Ref.Data*, **11**: Supplement No. 2.

VITA

Eric Martin Yezdimer was born on June 25, 1976 near Wilmington, Delaware. He attended public schools in New Castle County, Delaware, where he eventually graduated from John Dickinson High School in June 1994. He then enrolled at the University of Delaware and became heavily involved in undergraduate research, under the guidance of Professor Robert H. Wood. It was here that he first became interested in hydrothermal chemistry. In May of 1998 he received a Bachelor of Science degree in Chemistry from the University of Delaware. He then moved to Knoxville, Tennessee in pursuit of a graduate career in physical chemistry and began taking classes at the University of Tennessee during the fall of 1998. After supporting himself during his first year as a teaching assistant, he received a full research assistantship under the care of Professor Peter T. Cummings. Eager to expand his horizons beyond the academic world, he would later participate in a short three-month internship at DOW Chemical in Midland, Michigan during the summer of 2000. One year later he completed his PhD in physical chemistry from the University of Tennessee.